

# Measures of Residual Risk with Connections to Regression, Risk Tracking, Surrogate Models, and Ambiguity<sup>1</sup>

*R. Tyrrell Rockafellar*

Department of Mathematics  
University of Washington  
rtr@uw.edu

*Johannes O. Royset*

Operations Research Department  
Naval Postgraduate School  
joroyst@nps.edu

**Abstract.** Measures of residual risk are developed as extension of measures of risk. They view a random variable of interest in concert with an auxiliary random vector that helps to manage, predict, and mitigate the risk in the original variable. Residual risk can be exemplified as a quantification of the improved situation faced by a hedging investor compared to that of a single-asset investor, but the notion reaches further with deep connections emerging with forecasting and generalized regression. We establish the fundamental properties in this framework and show that measures of residual risk along with generalized regression can play central roles in the development of risk-tuned approximations of random variables, in tracking of statistics, and in estimation of the risk of conditional random variables. The paper ends with dual expressions for measures of residual risk, which lead to further insights and a new class of distributionally robust optimization models.

**Keywords:** risk measures, residual risk, generalized regression, surrogate estimation, optimization under stochastic ambiguity

**Date:** April 15, 2015

## 1 Introduction

Quantification of the “risk” associated with possible outcomes of a stochastic phenomenon, as described by a random variable, is central to much of operations research, economics, reliability engineering, and related areas. Measures of risk are important tools in this process that not only quantify risk, but also facilitate subsequent optimization of the parameters on which risk might depend; see for example the recent reviews [13, 26, 25]. In this paper, we extend the concept of risk measures to situations where the random variable of interest is viewed in concert with a related random vector that helps to manage, predict, and mitigate the risk in the original variable. A strategy of hedging in financial engineering, where the effect of potential losses from an investment is reduced by taking positions in correlated

---

<sup>1</sup>This material is based upon work supported in part by the U. S. Air Force Office of Scientific Research under FA9550-11-1-0206, F1ATAO1194GOO1, and F4FGA04094G003 as well as DARPA under HR0011412251.

<b>Report Documentation Page</b>			Form Approved OMB No. 0704-0188		
<p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>					
1. REPORT DATE <b>15 APR 2015</b>	2. REPORT TYPE	3. DATES COVERED <b>00-00-2015 to 00-00-2015</b>			
<b>4. TITLE AND SUBTITLE</b> <b>Measures of Residual Risk with Connections to Regression, Risk Tracking, Surrogate Models, and Ambiguity</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
<b>6. AUTHOR(S)</b>				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> <b>Naval Postgraduate School (NPS),Operations Research Department,Monterey,CA,93943</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> <b>Approved for public release; distribution unlimited</b>					
<b>13. SUPPLEMENTARY NOTES</b> <b>SIAM J. Optimization, to appear.</b>					
<b>14. ABSTRACT</b> <b>Measures of residual risk are developed as extension of measures of risk. They view a random variable of interest in concert with an auxiliary random vector that helps to manage, predict and mitigate the risk in the original variable. Residual risk can be exemplified as a quantification of the improved situation faced by a hedging investor compared to that of a single-asset investor, but the notion reaches further with deep connections emerging with forecasting and generalized regression. We establish the fundamental properties in this framework and show that measures of residual risk along with generalized regression can play central roles in the development of risk-tuned approximations of random variables, in tracking of statistics, and in estimation of the risk of conditional random variables. The paper ends with dual expressions for measures of residual risk, which lead to further insights and a new class of distributionally robust optimization models.</b>					
<b>15. SUBJECT TERMS</b>					
<b>16. SECURITY CLASSIFICATION OF:</b> a REPORT <b>unclassified</b>			<b>17. LIMITATION OF ABSTRACT</b> <b>Same as Report (SAR)</b>	<b>18. NUMBER OF PAGES</b> <b>34</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b>
b ABSTRACT <b>unclassified</b>	c THIS PAGE <b>unclassified</b>				

instruments, is a basic example that motivates our definition of *measures of residual risk*. However, measures of residual risk extend much beyond hedging and, in fact, lead to new measures of risk as well as deep-rooted connections with regression, risk-averse forecasting, and a multitude of applications.

For a random variable  $Y$  of primary interest and a related random vector  $X = (X_1, X_2, \dots, X_n)$ , we examine the situation where the goal is to find a *regression function*  $f$  such that  $Y$  is well approximated by  $f(X)$ . Presumably  $X$  is somehow more accessible than  $Y$ , making  $f(X)$  an attractive surrogate for  $Y$ . An example of such *surrogate estimation* arises in “factor models” in financial investment applications (see for example [6, 12]), where  $Y$  is the loss associated with a particular position and  $X$  a vector describing a small number of macroeconomic “factors” such as interest rates, inflation level, and GDP growth. In forecasting,  $f(X)$  might be the (random) forecast of the phenomenon described by  $Y$ , with its expectation  $E[f(X)]$  being an associated point prediction. In “uncertainty quantification” (see for example [14, 7]), one considers the output, described by a random variable  $Y$ , of a system subject to random input  $X$  whose distribution might be assumed known. Then, a regression function  $f$  leads to an accessible surrogate estimate  $f(X)$  of the unknown system output  $Y$ .

In surrogate estimation, traditionally, the focus has been on least-squares regression and its quantification of the difference between  $Y$  and  $f(X)$  in terms of mean squared error (MSE). In a risk-averse context where high realizations of  $Y$  are undesirable beyond any compensation by occasional low realizations, the symmetric view of errors inherent in MSE might be inappropriate and the consideration of *generalized, risk-averse regression* becomes paramount. A fundamental goal would then be, for a given measure of risk  $\mathcal{R}$ , to construct a regression function  $f$  such that

$$\mathcal{R}(Y) \leq \mathcal{R}(f(X)) + \text{possibly an error term.}$$

Initial work in this direction includes [22], which establishes such conservative surrogate estimates through generalized regression. We obtain the same result under weaker assumptions, develop means to assess the goodness-of-fit in generalized regression, examine the stability of regression functions, and make fundamental connections between such regression, surrogate estimation, and measures of residual risk.

Generalized regression also plays a central role in situations where the random vector  $X$ , at least eventually, comes under the control of a decision maker and the primary interest is then in the *conditional* random variable  $Y$  given  $X = x$ , which we denote by  $Y_x$ . For example, the goal might be to *track* a given statistic of  $Y_x$ , as it varies with  $x$ , or to minimize  $\mathcal{R}(Y_x)$  by choice of  $x$ , under a given measure of risk  $\mathcal{R}$ . The former situation is a theme of regression analysis, but we here go beyond expectations and quantiles, a traditional focus, and consider general classes of statistics. The latter situation is the standard setting of risk-averse stochastic programming; see for example [13, 26]. Due to incomplete distributional information about  $Y_x$  for every  $x$  as well as the computational cost of evaluating  $\mathcal{R}(Y_x)$  for numerous  $x$ , for example within an optimization algorithm, it might be beneficial in this situation to develop a regression function  $f$  such that

$$\text{for } x \text{ in a subset of interest, } \mathcal{R}(Y_x) \approx f(x).$$

Such a regression function provides an inexpensive substitute for  $\mathcal{R}(Y_x), x \in \mathbb{R}^n$ , within optimization

models. We refer to this situation as *risk tracking*, which in general cannot be carried out with precision; see [21] for a discussion in the context of superquantile/CVaR risk measures. Therefore, we look at *conservative risk tracking*, where  $f$  provides an (approximate) upper bound on  $\mathcal{R}(Y_x)$ ,  $x \in \mathbb{R}^n$ .

In the particular case of superquantile/CVaR risk measures, kernel-based estimators for the conditional probability density functions, integration, and inversion lead to estimates of conditional superquantiles [29, 4, 11]. Likewise, weighted-sums-of-conditional quantiles also give estimators of conditional superquantiles [20, 5, 15]. More generally, there is an extensive literature on estimating conditional distribution functions using nonparametric kernel estimators (see for example [9]) and transformation models (see for example [10]). Of course, with an estimate of a conditional distribution function, it is typically straightforward to estimate a statistic of  $Y_x$  and/or  $\mathcal{R}(Y_x)$  as parameterized by  $x$  for any law-invariant risk measure. However, it is generally difficult to obtain quality estimates of such conditional distribution functions and so here we focus on obtaining (conservative) estimates of statistics and risk directly.

It is well known through convex duality that many measures of risk quantify the risk in a random variable  $Y$  to be the worst-case expected value of  $Y$  over a risk envelope, often representing a set of alternative probability distributions; see for example [26] for a summary of results. We develop parallel, dual expressions for measures of residual risk and show that knowledge about a related random vector  $X$  leads to a *residual risk envelope* that is typically smaller than the original risk envelope. In fact,  $X$  gives rise to a new class of distributionally robust and computationally tractable optimization models that is placed between an expectation-minimization model and a distributionally robust model generated by a risk measure. The new models are closely allied with moment-matching of the related random vector  $X$ . Dual expressions of measures of residual risk through residual risk envelopes provide the key tool in this construction.

The contributions of the paper therefore lie in the introduction of measures of residual risk, the analysis of generalized regression, the discovery of the connections between residual risk and regression, and the application of these concepts in risk-tuned surrogate models, statistic and risk tracking, and distributionally robust optimization. In the process, we also improve and simplify prior results on the connections between risk measures and other quantifiers.

The paper continues in Section 2 with a review of basic concepts, definitions of measures of risk and related quantifiers, and a theorem about connections among such quantifiers under relaxed assumptions. Section 3 defines measures of residual risk, analyzes their properties, and makes connections with generalized regression. Sections 4 and 5 examine surrogate estimation and tracking, respectively. Section 6 discusses duality and distributionally robust formulations of optimization problems. An appendix supplements the paper with examples of risk measures and other quantifiers.

## 2 Preliminaries and Risk Quadrangle Connections

This section establishes terminology and provides connections among measures of risk and related quantities. We follow the *risk quadrangle framework* described in [26], but relax requirements in definitions and thereby extend the reach of that framework. We consider random variables defined on a probability

space  $(\Omega, \mathcal{F}, \mathbb{P})$  and restrict the attention to the subset  $\mathcal{L}^2 := \{Y : \Omega \rightarrow \mathbb{R} \mid Y \text{ measurable}, E[Y^2] < \infty\}$  of random variables with finite second moments. Although much of the discussion holds under weaker assumptions, among other issues we avoid technical complications related to paired topological spaces in duality statements under this restriction; see [28] for treatment of risk measures on more general spaces. We equip  $\mathcal{L}^2$  with the standard norm  $\|\cdot\|_2$  and convergence of random variables in  $\mathcal{L}^2$  will be in terms of the corresponding (strong) topology, if not specified otherwise. We adopt a perspective concerned about high values of random variables, which is natural in the case of “losses” and “costs.” A trivial sign change adjusts the framework to cases where low values, instead of high values, are undesirable.

We examine functionals  $\mathcal{F} : \mathcal{L}^2 \rightarrow (-\infty, \infty]$ , with measures of risk being specific instances. As we see below, several other functionals also play key roles. The following properties of such functionals arise in various combinations<sup>2</sup>:

- Constancy equivalence:  $\mathcal{F}(Y) = c_0$  for constant random variables  $Y \equiv c_0 \in \mathbb{R}$ .
- Convexity:  $\mathcal{F}((1 - \tau)Y + \tau Y') \leq (1 - \tau)\mathcal{F}(Y) + \tau\mathcal{F}(Y')$  for all  $Y, Y'$  and  $\tau \in (0, 1)$ .
- Closedness:  $\{Y \in \mathcal{L}^2 \mid \mathcal{F}(Y) \leq c_0\}$  is closed for all  $c_0 \in \mathbb{R}$ .
- Averseness:  $\mathcal{F}(Y) > E[Y]$  for nonconstant  $Y$ .
- Positive homogeneity:  $\mathcal{F}(\lambda Y) = \lambda\mathcal{F}(Y)$  and for every  $\lambda \geq 0$  and  $Y$ .
- Monotonicity:  $\mathcal{F}(Y) \leq \mathcal{F}(Y')$  when  $Y \leq Y'$ .
- Subadditivity:  $\mathcal{F}(Y + Y') \leq \mathcal{F}(Y) + \mathcal{F}(Y')$  for all  $Y, Y'$ .
- Finiteness:  $\mathcal{F}(Y) < \infty$  for all  $Y$ .

We note that convexity along with positive homogeneity is equivalent to subadditivity along with positive homogeneity. Closedness is also called lower semicontinuity.

Through conjugate duality (see [23] for a more general treatment), every closed convex functional  $\mathcal{F} : \mathcal{L}^2 \rightarrow (-\infty, \infty]$ ,  $\mathcal{F} \not\equiv \infty$ , is expressed by

$$\mathcal{F}(Y) = \sup_{Q \in \text{dom } \mathcal{F}^*} \left\{ E[QY] - \mathcal{F}^*(Q) \right\} \text{ for } Y \in \mathcal{L}^2, \quad (1)$$

where  $\mathcal{F}^* : \mathcal{L}^2 \rightarrow (-\infty, \infty]$  is the conjugate to  $\mathcal{F}$ , also a closed convex functional not identical to  $\infty$ , given by

$$\mathcal{F}^*(Q) = \sup_{Y \in \text{dom } \mathcal{F}} \left\{ E[QY] - \mathcal{F}(Y) \right\} \text{ for } Q \in \mathcal{L}^2, \quad (2)$$

and  $\text{dom } \mathcal{F}$  is the effective domain of  $\mathcal{F}$ , i.e.,  $\text{dom } \mathcal{F} := \{Y \in \mathcal{L}^2 \mid \mathcal{F}(Y) < \infty\}$ , and likewise for  $\text{dom } \mathcal{F}^*$ . Both  $\text{dom } \mathcal{F}$  and  $\text{dom } \mathcal{F}^*$  are necessarily nonempty and convex. The following facts about such functionals are used in the paper.  $\mathcal{F}$  is positively homogenous if and only if  $\mathcal{F}^*(Q) = 0$  for  $Q \in \text{dom } \mathcal{F}^*$ .  $\mathcal{F}$  is monotonic if and only if  $Q \geq 0$  for  $Q \in \text{dom } \mathcal{F}^*$ . The elements of the subdifferential  $\partial\mathcal{F}(Y) \subset \mathcal{L}^2$  for  $Y \in \mathcal{L}^2$  are those  $Q$  satisfying the subgradient inequality

$$\mathcal{F}(Y') \geq \mathcal{F}(Y) + E[Q(Y' - Y)] \text{ for all } Y' \in \mathcal{L}^2.$$

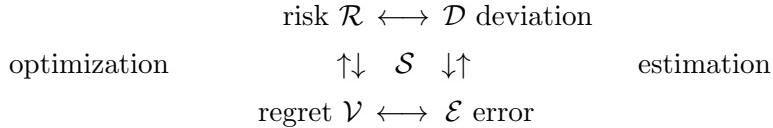
---

<sup>2</sup>Extended real-valued calculus is handled in the usually manner:  $0 \cdot \infty = 0$  and  $0 \cdot (-\infty) = 0$ ;  $a \cdot \infty = \infty$  and  $a \cdot (-\infty) = -\infty$  for  $a > 0$ ;  $\infty + \infty = \infty + (-\infty) = (-\infty) + \infty = \infty$ , and  $-\infty + (-\infty) = -\infty$ .

Moreover,  $\partial\mathcal{F}(Y) = \operatorname{argmax}_Q\{E[QY] - \mathcal{F}^*(Q)\}$  and this set is nonempty and weakly compact for all<sup>3</sup>  $Y \in \operatorname{int}(\operatorname{dom}\mathcal{F})$ .

We next turn the attention to specific functionals, referred to as measures of risk, regret, error, and deviation, that are tied together in quadrangles of risk with connections to risk optimization and statistical estimation; see Diagram 1 and the subsequent development.

---



**Diagram 1: The Fundamental Risk Quadrangle**

A *measure of risk* is a functional  $\mathcal{R}$  that assigns to a random variable  $Y \in \mathcal{L}^2$  a value  $\mathcal{R}(Y)$  in  $(-\infty, \infty]$  as a quantification of its risk. We give examples of measures of risk as well as other “measures” throughout the article and in the Appendix.

$\mathcal{R}$  is *regular* if it satisfies constancy equivalence, convexity, closedness, and averseness.

We observe that for a regular risk measure,  $\mathcal{R}(Y + c_0) = \mathcal{R}(Y) + c_0$  for any  $Y \in \mathcal{L}^2$  and  $c_0 \in \mathbb{R}$ ; see for example [26]. Regular measures of risk are related to, but distinct from coherent measures of risk [1] and convex risk functions [28]; see [26] for a discussion.

The effective domain  $\mathcal{Q} := \{Q \in \mathcal{L}^2 \mid \mathcal{R}^*(Q) < \infty\}$  of the conjugate  $\mathcal{R}^*$  to a regular measure of risk  $\mathcal{R}$  is called a *risk envelope*.

Consequently, maximization in (1) takes place over the risk envelope when  $\mathcal{F}$  is a regular measure of risk  $\mathcal{R}$ . Moreover,

a  $Q \in \mathcal{Q}$  that attains the supremum for  $Y \in \mathcal{L}^2$ , i.e.,  $\mathcal{R}(Y) = E[QY] - \mathcal{R}^*(Q)$ , is called a *risk identifier* at  $Y$  for  $\mathcal{R}$ , with all such  $Q$  forming the set  $\partial\mathcal{R}(Y)$ .

The nonemptiness of such subdifferentials ensures that there exists a risk identifier for all  $Y \in \operatorname{int}(\operatorname{dom}\mathcal{F})$ .

Closely connected to risk is the notion of regret, which in many ways is more fundamental. A *measure of regret* is a functional  $\mathcal{V}$  that assigns to a random variable  $Y \in \mathcal{L}^2$  a value  $\mathcal{V}(Y)$  in  $(-\infty, \infty]$  that quantifies the current displeasure with the mix of possible (future) outcomes for  $Y$ .

$\mathcal{V}$  is *regular* if it satisfies convexity and closedness as well as the property:

$$\mathcal{V}(0) = 0, \text{ but } \mathcal{V}(Y) > E[Y] \text{ when } Y \neq 0.$$

---

<sup>3</sup>We denote the (strong) topological interior of  $U \subset \mathcal{L}^2$  by  $\operatorname{int} U$ .

Regularity is here defined more broadly than in [26], where an additional condition is required. If  $Y$  is a financial loss, then  $\mathcal{V}(Y)$  can be interpreted as the monetary compensation demanded for assuming responsibility for covering the loss  $Y$ . We note that  $\mathcal{V}(Y)$  can be viewed simply as a reorientation of classical “utility” towards losses. Moreover, one can construct a regular measure of regret  $\mathcal{V}$  from a normalized concave utility function  $u : \mathbb{R} \rightarrow \mathbb{R}$ , with  $u(0) = 0$  and  $u(y) < y$  when  $y \neq 0$ , by setting  $\mathcal{V}(Y) = -E[u(-Y)]$ .

In regression, “error” plays the central role. A *measure of error* is a functional  $\mathcal{E}$  that assigns to a random variable  $Y \in \mathcal{L}^2$  a value  $\mathcal{E}(Y)$  in  $[0, \infty]$  that quantifies its nonzeroness.

$\mathcal{E}$  is *regular* if it satisfies convexity and closedness as well as the property:

$$\mathcal{E}(0) = 0, \text{ but } \mathcal{E}(Y) > 0 \text{ when } Y \neq 0.$$

Again, we define regularity more broadly than in [26]<sup>4</sup>.

An extension of the notion of standard deviation also emerges. A *measure of deviation* is a functional  $\mathcal{D}$  that assigns to a random variable  $Y \in \mathcal{L}^2$  a value  $\mathcal{D}(Y)$  in  $[0, \infty]$  that quantifies its nonconstancy.

$\mathcal{D}$  is *regular* if it satisfies convexity and closedness as well as the property:

$$\mathcal{D}(Y) = 0 \text{ for constant random variables } Y \equiv c_0 \in \mathbb{R}, \text{ but } \mathcal{D}(Y) > 0 \text{ for nonconstant } Y \in \mathcal{L}^2.$$

Error minimization is the focus of regression. In the case of an error measure  $\mathcal{E}$ , the *statistic*

$$\mathcal{S}(Y) := \operatorname{argmin}_{c_0 \in \mathbb{R}} \mathcal{E}(Y - c_0) \tag{3}$$

is the quantity obtained through such minimization. It is the set of scalars, in many cases a singleton, that best approximate  $Y$  in the sense of error measure  $\mathcal{E}$ . We refer to the Appendix for examples of measures of risk, regret, error, and deviation, and corresponding statistics.

Before giving connections among the various measures and statistics, we establish the following technical result. The proof is a specialization of the argument in the proof of Lemma 3.3 provided below and is therefore omitted.

**2.1 Lemma** *For a regular measure of error  $\mathcal{E}$  and sequence  $\{c_0^\nu\}_{\nu=1}^\infty$  of scalars, the following holds: If  $Y^\nu \in \mathcal{L}^2$  and  $b^\nu \in \mathbb{R}$  converge to  $Y \in \mathcal{L}^2$  and  $b \in \mathbb{R}$ , respectively, and  $\mathcal{E}(Y^\nu - c_0^\nu) \leq b^\nu$  for all  $\nu$ , then  $\{c_0^\nu\}_{\nu=1}^\infty$  is bounded and any accumulation point  $c_0$  satisfies  $\mathcal{E}(Y - c_0) \leq b$ .*

Connections among regular measures and statistics are given by the following results, which extend the Quadrangle Theorem in [26] to the broader class of regular measures defined here and also include additional characterizations of deviation measures and statistics.

**2.2 Theorem** (risk quadrangle connections) *Regular measures of risk, regret, deviation, and error are related as follows:*

---

<sup>4</sup>The extra conditions, on the behavior of certain limits, have turned out to be superfluous for the results in [26].

(i) The relations

$$\mathcal{R}(Y) = \mathcal{D}(Y) + E[Y] \text{ and } \mathcal{D}(Y) = \mathcal{R}(Y) - E[Y] \quad (4)$$

give a one-to-one correspondence between regular measures or risk  $\mathcal{R}$  and regular measures of deviation  $\mathcal{D}$ . Here,  $\mathcal{R}$  is positively homogeneous if and only if  $\mathcal{D}$  is positively homogeneous. Moreover,  $\mathcal{R}$  is monotonic if and only if  $\mathcal{D}(Y) \leq \sup Y - E[Y]$  for all  $Y \in \mathcal{L}^2$ .

(ii) The relations

$$\mathcal{V}(Y) = \mathcal{E}(Y) + E[Y] \text{ and } \mathcal{E}(Y) = \mathcal{V}(Y) - E[Y] \quad (5)$$

give a one-to-one correspondence between regular measures of regret  $\mathcal{V}$  and regular measures of error  $\mathcal{E}$ . Here,  $\mathcal{V}$  is positively homogeneous if and only if  $\mathcal{E}$  is positively homogeneous. Moreover,  $\mathcal{V}$  is monotonic if and only if  $\mathcal{E}(Y) \leq |E[Y]|$  for all  $Y \leq 0$ .

(iii) For any regular measure of regret  $\mathcal{V}$ , a regular measure of risk is obtained by

$$\mathcal{R}(Y) = \min_{c_0 \in R} \left\{ c_0 + \mathcal{V}(Y - c_0) \right\}. \quad (6)$$

If  $\mathcal{V}$  is positively homogeneous, then  $\mathcal{R}$  is positively homogeneous. If  $\mathcal{V}$  is monotonic, then  $\mathcal{R}$  is monotonic.

(iv) For any regular measure of error  $\mathcal{E}$ , a regular measure of deviation is obtained by

$$\mathcal{D}(Y) = \min_{c_0 \in R} \mathcal{E}(Y - c_0). \quad (7)$$

If  $\mathcal{E}$  is positively homogeneous, then  $\mathcal{D}$  is positively homogeneous. If  $\mathcal{E}$  satisfies  $\mathcal{E}(Y) \leq |E[Y]|$  for all  $Y \leq 0$ , then  $\mathcal{D}$  satisfies  $\mathcal{D}(Y) \leq \sup Y - E[Y]$  for all  $Y \in \mathcal{L}^2$ . Moreover,  $\mathcal{D}(Y + c_0) = \mathcal{D}(Y)$  for any  $Y \in \mathcal{L}^2$  and  $c_0 \in \mathbb{R}$ .

(v) For corresponding  $\mathcal{V}$  and  $\mathcal{E}$  according to (ii) and  $Y \in \mathcal{L}^2$ , the statistic

$$\mathcal{S}(Y) = \operatorname{argmin}_{c_0 \in R} \mathcal{E}(Y - c_0) = \operatorname{argmin}_{c_0 \in R} \left\{ c_0 + \mathcal{V}(Y - c_0) \right\}. \quad (8)$$

It is a nonempty closed bounded interval as long as  $\mathcal{V}(Y - c_0)$ , or equivalently  $\mathcal{E}(Y - c_0)$ , is finite for some  $c_0 \in \mathbb{R}$ . Moreover,  $\mathcal{S}(Y + c_0) = \mathcal{S}(Y) + \{c_0\}$  for any  $Y \in \mathcal{L}^2$  and  $c_0 \in \mathbb{R}$ , and  $\mathcal{S}(0) = \{0\}$ .

**Proof.** Part (i) is a direct consequence of the regularity of  $\mathcal{R}$  and  $\mathcal{D}$ , which are unchanged from the Quadrangle Theorem in [26].

Part (ii) is also a direct consequence of the regularity of  $\mathcal{V}$  and  $\mathcal{E}$ , and the broadening, compared to [26], of the class of regular measures does not require modified arguments.

The claims in Part (iii) about positive homogeneity and monotonicity follow easily and by the same arguments as those leading to the same conclusions in [26]. However, the claims that the infimum in (6) is attained and indeed produces a regular measure of risk require a new argument. Since

$$c_0 + \mathcal{V}(Y - c_0) = \mathcal{E}(Y - c_0) + E[Y]$$

by Part (ii), it suffices to consider minimization of  $\mathcal{E}(Y - c_0)$ . First, suppose that  $\inf_{c_0} \mathcal{E}(Y - c_0) < \infty$ . Then, there exist  $\{c_0^\nu\}_{\nu=1}^\infty$  and  $\{\varepsilon^\nu\}_{\nu=1}^\infty$  such that  $\varepsilon^\nu \rightarrow 0$  and

$$\mathcal{E}(Y - c_0^\nu) \leq \inf_{c_0 \in \mathcal{R}} \mathcal{E}(Y - c_0) + \varepsilon^\nu \text{ for all } \nu.$$

Applying Lemma 2.1 with  $Y^\nu = Y$ ,  $b^\nu = \inf_{c_0 \in \mathcal{R}} \mathcal{E}(Y - c_0) + \varepsilon^\nu$ , and  $b = \inf_{c_0 \in \mathcal{R}} \mathcal{E}(Y - c_0)$ , we obtain that  $\{c_0^\nu\}_{\nu=1}^\infty$  is bounded, that there exists a scalar  $c_0^*$  and a subsequence  $\{c_0^\nu\}_{\nu \in \mathcal{N}}$ , with  $c_0^\nu \xrightarrow{\mathcal{N}} c_0^*$ , and that

$$\mathcal{E}(Y - c_0^*) \leq \inf_{c_0 \in \mathcal{R}} \mathcal{E}(Y - c_0).$$

Consequently,  $c_0^* \in \operatorname{argmin}_{c_0} \mathcal{E}(Y - c_0)$ . Second, if  $\inf_{c_0} \mathcal{E}(Y - c_0) = \infty$ , then  $\mathcal{R} = \operatorname{argmin}_{c_0} \mathcal{E}(Y - c_0)$ . Thus, the infimum in (6) is attained in both cases. Next, we consider closedness. Suppose that  $Y^\nu \rightarrow Y$ ,  $c_0^\nu \in \operatorname{argmin}_{c_0} \mathcal{E}(Y^\nu - c_0)$ , and  $\mathcal{E}(Y^\nu - c_0^\nu) \leq b \in \mathcal{R}$  for all  $\nu$ . Hence,  $\mathcal{R}(Y^\nu) - E[Y^\nu] = \mathcal{E}(Y^\nu - c_0^\nu) \leq b$  for all  $\nu$ . An application of Lemma 2.1 implies that there exists a scalar  $c_0^*$  and a subsequence  $\{c_0^\nu\}_{\nu \in \mathcal{N}}$ , with  $c_0^\nu \xrightarrow{\mathcal{N}} c_0^*$ , and  $\mathcal{E}(Y - c_0^*) \leq b$ . Consequently,  $\mathcal{R}(Y) - E[Y] = \min_{c_0} \mathcal{E}(Y - c_0) \leq \mathcal{E}(Y - c_0^*) \leq b$ , which establishes the closedness of  $\mathcal{R}(\cdot) - E[\cdot]$ . The expectation functional is finite and continuous on  $\mathcal{L}^2$  so the closedness of  $\mathcal{R}$  is also established. Since constancy equivalence, convexity, and averseness follow trivially,  $\mathcal{R}$  is regular.

Part (iv) follows from Parts (i)-(iii), with the exception of the last claim, which is a consequence of the fact that  $\mathcal{R}(Y + c_0) = \mathcal{R}(Y) + c_0$  for regular measures of risk.

In Part (v), the alternative expression for  $\mathcal{S}(Y)$  follows by Part (ii). The closedness and convexity of  $\mathcal{S}(Y)$  are obvious from the closedness and convexity of  $\mathcal{E}$ . Its nonemptiness is a consequence of the proof of Part (ii). An application of Lemma 2.1, with  $Y^\nu = Y$ ,  $b^\nu = b = \mathcal{D}(Y)$ , and  $c_0^\nu \in \mathcal{S}(Y)$ , establishes the boundedness of  $\mathcal{S}$ . The calculus rules for  $\mathcal{S}$  follow trivially from the definition of the statistic.  $\square$

Regular measures of risk, regret, error, and deviation as well as statistics related according to Theorem 2.2 are said to be in *correspondence*. In contexts where  $Y$  is a monetary loss, then the scalar  $c_0$  in (6) can be interpreted as the investment today in a risk-free asset that minimizes the displeasure associated with taking responsibility of a future loss  $Y$ . Even in the absence of a risk-free investment opportunity,  $c_0$  could represent a certain future expenditure that allows one to offset the loss  $Y$ . In other contexts where one aims to forecast a realization of  $Y$ ,  $c_0 \in \mathcal{S}(Y)$  can be viewed as a point forecast of that realization and (6) as a tradeoff between making a low point forecast and the displeasure derived from making an “incorrect” forecast. We provide further interpretations in the next section as we extend the notion of risk measure.

### 3 Residual Measures of Risk

A measure of risk applies to a single random variable. However, in many contexts the scope needs to be widened by also looking at other related random variables that hopefully might provide insight, improve prediction, and reduce “risk.”

In this section, we introduce a *measure of residual risk* that extends a measure of risk to a context involving not only a random variable  $Y$ , still of primary interest, but also a related random vector  $X = (X_1, \dots, X_n) \in \mathcal{L}_n^2 := \mathcal{L}^2 \times \dots \times \mathcal{L}^2$ . The definition is motivated by tradeoffs experienced by forecasters and investors, but as we shall see connections with regression, surrogate models, and distributional robustness are also profound. We start with the definition and motivations, and proceed to fundamental properties and connections with generalized regression.

### 3.1 Definition and Motivation

As an extension of the trade-off formula (6) for a measure of risk, we adopt the following definition of a measure of residual risk.

**3.1 Definition** (measures of residual risk) *For given  $X \in \mathcal{L}_n^2$  and regular measure of regret  $\mathcal{V}$ , we define the associated measure of residual risk (in the context of affine approximating functions) to be the functional  $\mathcal{R}(\cdot|X) : \mathcal{L}^2 \rightarrow [-\infty, \infty]$  given by*

$$\mathcal{R}(Y|X) := \inf \left\{ E[f(X)] + \mathcal{V}(Y - f(X)) \mid f \text{ affine} \right\} \text{ for } Y \in \mathcal{L}^2. \quad (9)$$

The quantity  $\mathcal{R}(Y|X)$  is the residual risk of  $Y$  with respect to  $X$  that comes from  $\mathcal{V}$ .

We observe that since  $\mathcal{L}^2$  is a linear space,  $Y - f(X) \in \mathcal{L}^2$  when  $f$  is affine. Consequently,  $\mathcal{R}(\cdot|X)$  is well defined. Two examples motivate the definition:

**Example 1: Prediction.** Consider a situation where we would like predict the peak electricity demand in a region for tomorrow. Today this quantity is unknown and we can think of it as a random variable  $Y$ . To help us make the prediction, temperature, dew point, and cloud cover forecast for tomorrow are available, possibly for different hours of the day. Suppose that the forecast gives the joint probability distribution for these quantities viewed as a random vector  $X$  and that our (random) prediction of tomorrow's electricity demand is of the form  $f(X)$ , with  $f$  an affine function. Our point forecast is  $E[f(X)]$ . The point forecast will be used to support decisions about power generation, where higher peak demand causes additional costs and challenges, and we therefore prefer to select  $f$  such that  $E[f(X)]$  is as low as possible. Of course, we need to balance this with the need to avoid underpredicting the demand. Suppose that a regular measure of regret  $\mathcal{V}$  quantifies our displeasure with under- and overprediction. Specifically,  $\mathcal{V}(Y - f(X))$  is the regret associated with  $f$ . For example, if  $\mathcal{V} = E[\max\{\cdot, 0\}]/(1 - \alpha)$ ,  $\alpha \in (0, 1)$ , then we are indifferent to overpredictions and feel increasing displeasure from successively larger underpredictions. A possible approach to constructing  $f$  would be to use historical data about peak demand, temperature, dew point, and cloud cover to find an affine function  $f$  such that both  $E[f(X)]$  and  $\mathcal{V}(Y - f(X))$  are low when  $(X, Y)$  is assumed to follow the empirical distribution given by the data. This bi-objective optimization problem is solved in (9) through scalarization with equal weights between the objectives. (Other weights simply indicate another choice of  $\mathcal{V}$ .) The resulting optimal value is the residual risk of  $Y$  with respect to  $X$  and consists of the point forecast plus a “premium” quantifying our displeasure with an “incorrect” forecast. In contrast, if  $f$  is

restricted to the constant functions, then (9) reduces to (6) and no information about  $X$  is included. Specifically, historical data about peak demand is used to find a constant  $c_0$  that minimizes (6), i.e., makes both the point forecast  $c_0$  and the regret  $\mathcal{V}(Y - c_0)$  low. The optimal value is the risk of  $Y$ , which again consists of a point forecast plus a premium quantifying our displeasure with “getting it wrong.” A high value of risk or residual risk therefore implies that we are faced with an unpleasant situation where the forecast for the peak demand as well as our regret about the forecast are relatively high. The contributions from each term are easily determined in the process of solving (6) and (9). The restriction to constant functions  $f$  clearly shows that

$$\mathcal{R}(Y|X) \leq \mathcal{R}(Y).$$

Consequently, the situation can only improve as one brings in information about temperature, dew point, and cloud cover and compute the forecast  $f(X)$  instead of  $c_0$ . Typically, the point forecast  $E[f(X)]$  will be lower than  $c_0$  and the associate regret  $\mathcal{V}(Y - f(X))$  will be lower than  $\mathcal{V}(Y - c_0)$ , at least the sum of point forecast and regret will not worsen when additional information is brought in. A quantification of the improvement is the difference between risk and residual risk. Of course, there is nothing special about electricity demand and many other situations can be viewed similarly.

It is possible to consider alternatives to the expectation-based “point-forecast”  $E[f(X)]$ , but a discussion of that subject carries us beyond the scope of the present paper. In the following, we write affine functions on  $\mathbb{R}^n$  in the form  $c_0 + \langle c, \cdot \rangle$  for  $c_0 \in \mathbb{R}$  and  $c \in \mathbb{R}^n$ , where the inner product  $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ . Consequently, for  $X \in \mathcal{L}_n^2$ ,  $f(X) = c_0 + \langle c, X \rangle$  is therefore a pointwise equality between random variables, i.e.,  $c_0 + \langle c, X \rangle$  is a random variable, say,  $Z$  given by  $Z(\omega) = c_0 + \langle c, X(\omega) \rangle$ ,  $\omega \in \Omega$ . An interpretation of residual risk arises also in a financial context:

**Example 2: Hedging investor.** Consider a loss  $Y$ , given in present money, that an individual faces at a future point in time. If the individual is passive, i.e., does not consider investment options that might potentially offset a loss, she might simply assess this loss according its regret  $\mathcal{V}(Y)$ , where  $\mathcal{V}$  is a regular measure of regret that quantifies the investor’s displeasure with the mix of possible losses. In view of the earlier comment about connections between regret and utility, this quantification is therefore quite standard and often used when comparing various alternative losses and gains. If the individual is more active and invests  $c_0 \in \mathbb{R}$  in a risk-free asset now, then the future regret, as perceived now, is reduced from  $\mathcal{V}(Y)$  to  $\mathcal{V}(Y - c_0)$  as  $c_0$  will be available at the future point in time to offset the loss  $Y$ . Though, the upfront cost  $c_0$  needs also to be considered, and the goal becomes to select the risk-free investment  $c_0$  such that  $c_0 + \mathcal{V}(Y - c_0)$  is minimized. According to (6), the resulting value is the corresponding risk  $\mathcal{R}(Y)$  and every  $c_0 \in \mathcal{S}(Y)$ , the corresponding statistic, furnishes the amount to be invested in the risk-free asset. To further mitigate the loss, the individual might consider purchasing  $c_i$  shares in a stock  $i$  with random value  $X_i$ , in present terms, at the future point in time. The price of each share is  $p_i = E[X_i]$ . Let  $i = 1, 2, \dots, n$ ,  $c = (c_1, \dots, c_n)$ ,  $p = (p_1, \dots, p_n)$ , and  $X = (X_1, \dots, X_n)$ . Then, since  $Y - [c_0 + \langle c, X \rangle]$  is the future hedged loss in present terms, the future regret, as perceived now, is reduced from  $\mathcal{V}(Y)$  to  $\mathcal{V}(Y - [c_0 + \langle c, X \rangle])$ . Though, the upfront cost  $c_0 + \langle c, p \rangle$  needs also to be

considered, and the goal becomes to select the risk-free investment  $c_0$  and the risky investments  $c \in \mathbb{R}^n$  that

$$\text{minimize } \left\{ c_0 + \langle c, p \rangle + \mathcal{V}(Y - [c_0 + \langle c, X \rangle]) \right\},$$

which according to (6) is equivalent to selecting the risky investments  $c \in \mathbb{R}^n$  that

$$\text{minimize } \left\{ \langle c, p \rangle + \mathcal{R}(Y - \langle c, X \rangle) \right\}.$$

The optimal values of these problems are the residual risk  $\mathcal{R}(Y|X)$ . The possibly nonoptimal choices of setting  $c_0 = 0$  and/or  $c = 0$  correspond to forfeiting moderation of the future loss through risk-free and/or risky investments and give the values  $\mathcal{R}(Y)$  and  $\mathcal{V}(Y)$ . Consequently,

$$\mathcal{R}(Y|X) \leq \mathcal{R}(Y) \leq \mathcal{V}(Y).$$

The differences between these quantities reflect the degree of benefit an investor derives by departing from the passive strategy of  $c_0 = 0$  and  $c = 0$  to various degrees. Of course, the ability to reduce risk by taking positions in the stocks is determined by the dependence between  $Y$  and  $X$ . In a decision making situation, when comparing two candidate random variables  $Y$  and  $Y'$ , an individual's preference of one over the other heavily depends on whether the comparison is carried out at the level of regret, i.e.,  $\mathcal{V}(Y)$  versus  $\mathcal{V}(Y')$ , as in the case of traditional expected utility theory, at the level of risk, i.e.,  $\mathcal{R}(Y)$  versus  $\mathcal{R}(Y')$ , as in the case of much of modern risk analysis in finance, or at the level of residual risk  $\mathcal{R}(Y|X)$  versus  $\mathcal{R}(Y'|X)$ . The latter perspective might provide a more comprehensive picture of the "risk" faced by the decision maker as it accounts for the opportunities that might exist to offset losses. The focus on residual risk in decision making is related to the extensive literature on real options (see for example [8] and references therein), where also losses and gains are viewed in concert with other decisions.

### 3.2 Basic Properties

We continue in this subsection by examining the properties of measures of residual risk. We often require the nondegeneracy of the auxiliary random vector  $X$ , which is defined as follows:

**3.2 Definition** (nondegeneracy) *We will say that an  $n$ -dimensional random vector  $X = (X_1, X_2, \dots, X_n) \in \mathcal{L}_n^2$  is nondegenerate if*

$$\langle c, X \rangle \text{ is a constant} \implies c = 0 \in \mathbb{R}^n.$$

We note that nondegeneracy is equivalent to linear independence of  $1, X_1, X_2, \dots, X_n$  as elements of  $\mathcal{L}^2$ . For  $X \in \mathcal{L}_n^2$ , we also define the subspace

$$\mathcal{Y}(X) := \{Y \in \mathcal{L}^2 \mid Y = c_0 + \langle c, X \rangle, c_0 \in \mathbb{R}, c \in \mathbb{R}^n\}.$$

Before giving the main properties, we establish the following technical result which covers and extends Lemma 2.1.

**3.3 Lemma** For a regular measure of error  $\mathcal{E}$  and sequence  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$ , with  $c_0^\nu \in \mathbb{R}$  and  $c^\nu \in \mathbb{R}^n$  for all  $\nu$ , the following holds:

If  $Y^\nu \in \mathcal{L}^2$ ,  $X^\nu \in \mathcal{L}_n^2$ , and  $b^\nu \in \mathbb{R}$  converge to  $Y \in \mathcal{L}^2$ ,  $X \in \mathcal{L}_n^2$ , and  $b \in \mathbb{R}$ , respectively,  $X$  is nondegenerate, and  $\mathcal{E}(Y^\nu - [c_0^\nu + \langle c^\nu, X^\nu \rangle]) \leq b^\nu$  for all  $\nu$ , then  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  is bounded and any accumulation point  $(c_0, c)$  satisfies  $\mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \leq b$ .

**Proof.** For the sake of a contradiction suppose that  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  is not bounded. Then, there exists a subsequence  $\{(c_0^\nu, c^\nu)\}_{\nu \in \mathcal{N}}$  such that  $\|(c_0^\nu, c^\nu)\| > 1$  for all  $\nu \in \mathcal{N}$ ,  $\|(c_0^\nu, c^\nu)\| \rightarrow^{\mathcal{N}} \infty$ , and  $(c_0^\nu, c^\nu)/\|(c_0^\nu, c^\nu)\| \rightarrow^{\mathcal{N}} (a_0, a) \neq 0$ , with  $a_0 \in \mathbb{R}$  and  $a \in \mathbb{R}^n$ . Let  $\lambda^\nu = 1/\|(c_0^\nu, c^\nu)\|$ . Since  $\mathcal{E}$  is convex and  $\mathcal{E}(0) = 0$ , we have that

$$\mathcal{E}(\lambda Y) \leq \lambda \mathcal{E}(Y) \text{ for } Y \in \mathcal{L}^2 \text{ and } \lambda \in [0, 1].$$

Consequently, for  $\nu \in \mathcal{N}$ ,

$$\lambda^\nu b^\nu \geq \lambda^\nu \mathcal{E}(Y^\nu - [c_0^\nu + \langle c^\nu, X^\nu \rangle]) \geq \mathcal{E}(\lambda^\nu Y^\nu - [\lambda^\nu c_0^\nu + \langle \lambda^\nu c^\nu, X^\nu \rangle]) \geq 0.$$

Since  $\lambda^\nu \rightarrow^{\mathcal{N}} 0$ ,  $\lambda^\nu b^\nu \rightarrow^{\mathcal{N}} 0$  and  $\lambda^\nu Y^\nu - [\lambda^\nu c_0^\nu + \langle \lambda^\nu c^\nu, X^\nu \rangle] \rightarrow^{\mathcal{N}} -[a_0 + \langle a, X \rangle]$ . These facts together with the closedness of  $\mathcal{E}$  imply that  $\mathcal{E}(-[a_0 + \langle a, X \rangle]) = 0$  and therefore also that  $a_0 + \langle a, X \rangle = 0$ . Since  $X$  is nondegenerate, this implies that  $a = 0$ . Then, however,  $a_0 = 0$ , and  $(a_0, a) = 0$ , which is a contradiction. Thus,  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  is bounded. The inequality  $\mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \leq b$  follows directly from the closedness of  $\mathcal{E}$ .  $\square$

Fundamental properties of measures of residual risk are given next.

**3.4 Theorem** (residual-risk properties) For given  $X \in \mathcal{L}_n^2$  and regular measures of regret  $\mathcal{V}$ , risk  $\mathcal{R}$ , deviation  $\mathcal{D}$ , and error  $\mathcal{E}$  in correspondence, the following facts about the associated measure of residual risk  $\mathcal{R}(\cdot|X)$  hold:

(i)  $\mathcal{R}(Y|X)$  satisfies the alternative formulae

$$\begin{aligned} \mathcal{R}(Y|X) &= \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle) \right\} \\ &= E[Y] + \inf_{c \in \mathbb{R}^n} \mathcal{D}(Y - \langle c, X \rangle) \\ &= E[Y] + \inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]). \end{aligned}$$

(ii)  $E[Y] \leq \mathcal{R}(Y|X) \leq \mathcal{R}(Y) \leq \mathcal{V}(Y)$  for all  $Y \in \mathcal{L}^2$ .

(iii)  $\mathcal{R}(\cdot|X)$  is convex and satisfies the constant equivalence property.

(iv) If  $\mathcal{V}$  is positively homogeneous, then  $\mathcal{R}(\cdot|X)$  is positively homogeneous. If  $\mathcal{V}$  is monotonic, then  $\mathcal{R}(\cdot|X)$  is monotonic.

(v) If  $X$  is a constant random vector, then  $\mathcal{R}(Y|X) = \mathcal{R}(Y)$ .

(vi) If  $X$  is nondegenerate, then  $\mathcal{R}(\cdot|X)$  is closed and the infimum in its definition as well as in the alternative formulae in (i) is attained.

(vii)  $\mathcal{R}(Y|X) = E[Y]$  if  $Y \in \mathcal{Y}(X)$ , whereas  $\mathcal{R}(Y|X) > E[Y]$  if  $Y \notin \mathcal{Y}(X)$  and  $X$  is nondegenerate.

**Proof.** Part (i) is a direct consequence of the relationships between corresponding measures given in Theorem 2.2. The first inequality in Part (ii) is a consequence of the fact that  $\mathcal{V} \geq E[\cdot]$  on  $\mathcal{L}^2$ . The second inequality follows by selecting the possibly nonoptimal solution  $c = 0$  in the first alternative formula and the third inequality by selecting  $c_0 = 0$  and  $c = 0$  in the definition

$$\mathcal{R}(Y|X) = \inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \left\{ c_0 + \langle c, E[X] \rangle + \mathcal{V}(Y - [c_0 + \langle c, X \rangle]) \right\}.$$

Part (v) is obtained from the first alternative formula in Part (i) and the fact that  $\mathcal{R}(Y+k) = \mathcal{R}(Y) + k$  for any  $k \in \mathbb{R}$ .

For Part (iii), convexity follows since the function  $(c_0, c, Y) \mapsto c_0 + \langle c, E[X] \rangle + \mathcal{V}(Y - [c_0 + \langle c, X \rangle])$  is convex on  $\mathbb{R} \times \mathbb{R}^n \times \mathcal{L}^2$ ; see for example [23, Theorem 1]. Constant equivalence is a consequence of Part (ii) and the fact that  $c_0 = E[Y] \leq \mathcal{R}(Y|X) \leq \mathcal{R}(Y) = c_0$  when  $Y \equiv c_0$ .

Part (iv) follows trivially from the definitions of positive homogeneity and monotonicity and Part (v) is likewise straightforwardly obtained.

Next we address Part (vi). First, we consider the minimization of  $\mathcal{E}(Y - [c_0 + \langle c, X \rangle])$ . Suppose that  $\inf_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) < \infty$ . Then, there exist  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$ , with  $c_0^\nu \in \mathbb{R}$  and  $c^\nu \in \mathbb{R}^n$ , as well as  $\{\varepsilon^\nu\}_{\nu=1}^\infty$  such that  $\varepsilon^\nu \rightarrow 0$  and

$$\mathcal{E}(Y - [c_0^\nu + \langle c^\nu, X \rangle]) \leq \inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) + \varepsilon^\nu \text{ for all } \nu.$$

Applying Lemma 3.3 with  $Y^\nu = Y$ ,  $X^\nu = X$ ,  $b^\nu = \inf_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) + \varepsilon^\nu$ , and  $b = \inf_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])$ , we obtain that  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  is bounded, that there exist  $c_0^* \in \mathbb{R}$ ,  $c^* \in \mathbb{R}^n$ , and a subsequence  $\{(c_0^\nu, c^\nu)\}_{\nu \in \mathcal{N}}$ , with  $(c_0^\nu, c^\nu) \xrightarrow{\mathcal{N}} (c_0^*, c^*)$ , and that

$$\mathcal{E}(Y - [c_0^* + \langle c^*, X \rangle]) \leq \inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]).$$

Consequently,  $(c_0^*, c^*) \in \operatorname{argmin}_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])$ . If  $\inf_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) = \infty$ , then  $\mathbb{R}^{n+1} = \operatorname{argmin}_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])$ . Thus, the error minimization in Part (i) is attained when  $X$  is nondegenerate. In view of (5), the infimum in the definition of residual risk is also attained. A nearly identical argument shows that the infima in the alternative formulae in (i) are also attained. Second, we consider closedness. Suppose that  $Y^\nu \rightarrow Y$ ,  $(c_0^\nu, c^\nu) \in \operatorname{argmin}_{c_0, c} \mathcal{E}(Y^\nu - [c_0 + \langle c, X \rangle])$ , and  $\mathcal{E}(Y^\nu - [c_0^\nu + \langle c, X \rangle]) \leq b \in \mathbb{R}$  for all  $\nu$ . Hence,  $\mathcal{R}(Y^\nu|X) - E[Y^\nu] = \mathcal{E}(Y^\nu - [c_0^\nu + \langle c^\nu, X \rangle]) \leq b$  for all  $\nu$ . An application of Lemma 3.3 implies that there exist  $c_0^* \in \mathbb{R}$ ,  $c^* \in \mathbb{R}^n$ , and a subsequence  $\{(c_0^\nu, c^\nu)\}_{\nu \in \mathcal{N}}$ , with  $(c_0^\nu, c^\nu) \xrightarrow{\mathcal{N}} (c_0^*, c^*)$ , and  $\mathcal{E}(Y - [c_0^* + \langle c^*, X \rangle]) \leq b$ . Consequently,  $\mathcal{R}(Y|X) - E[Y] = \min_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \leq \mathcal{E}(Y - [c_0^* + \langle c^*, X \rangle]) \leq b$ , which establishes the closedness of  $\mathcal{R}(\cdot|X) - E[\cdot]$ . The expectation functional is finite and continuous on  $\mathcal{L}^2$  so the closedness of  $\mathcal{R}(\cdot|X)$  is also established.

Finally, we consider Part (vii). Suppose that  $Y \in \mathcal{Y}(X)$ . Then, there exists  $\hat{c}_0 \in \mathbb{R}$  and  $\hat{c} \in \mathbb{R}^n$  such that  $Y = \hat{c}_0 + \langle \hat{c}, X \rangle$ . In view of Parts (i) and (ii)

$$\begin{aligned} E[Y] &\leq \mathcal{R}(Y|X) = \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle) \right\} \\ &\leq \langle \hat{c}, E[X] \rangle + \mathcal{R}(Y - \langle \hat{c}, X \rangle) \\ &= \langle \hat{c}, E[X] \rangle + \mathcal{R}(\hat{c}_0) \\ &= \hat{c}_0 + \langle \hat{c}, E[X] \rangle = E[Y], \end{aligned}$$

which establishes the first claim. Suppose that  $Y \notin \mathcal{Y}(X)$ . Then,  $Y - \langle c, X \rangle \neq c_0$  for any  $c_0 \in \mathbb{R}$  and  $c \in \mathbb{R}^n$ . Consequently,  $Y - \langle c, X \rangle$  is not a constant for any  $c \in \mathbb{R}^n$ , which by the averseness of  $\mathcal{R}$  implies that  $\mathcal{R}(Y - \langle c, X \rangle) > E[Y - \langle c, X \rangle]$ . If  $X$  is nondegenerate, then by Part (vi) there exists  $\bar{c} \in \mathbb{R}^n$  such that

$$\begin{aligned} \mathcal{R}(Y|X) &= \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle) \right\} \\ &= \langle \bar{c}, E[X] \rangle + \mathcal{R}(Y - \langle \bar{c}, X \rangle) \\ &> \langle \bar{c}, E[X] \rangle + E[Y - \langle \bar{c}, X \rangle] = E[Y], \end{aligned}$$

which completes the proof.  $\square$

We see from Theorem 3.4(i) that a measure of residual risk decomposes into an “irreducible” value  $E[Y]$  and a quantification of “nonzeroness” by an error measure of the difference between  $Y$  and an affine model in terms of  $X$ , that is reduced as much as possible by choosing  $c_0, c$  optimally.

A fundamental consequence of Theorem 3.4 is that for a nondegenerate  $X$ ,

a measure of residual risk is also a closed, convex, and constancy equivalent measure of risk.

The constructed risk measure is positively homogeneous if the underlying risk measure is positively homogeneous. Monotonicity is likewise inherited. When  $X$  is nondegenerate, it is also averse outside  $\mathcal{Y}(X)$ .

Further insight is revealed by the following trivial but informative example.

**Example 3: Normal random variables.** Suppose that  $X$  and  $Y$  are normal random variables with mean values  $\mu_X$  and  $\mu_Y$ , respectively, and standard deviations  $\sigma_X > 0$  and  $\sigma_Y$ , respectively. We here temporarily let  $X$  be scalar valued. Let  $\rho \in [-1, 1]$  be the correlation coefficient between  $X$  and  $Y$ , and  $G_Y(\alpha)$  be the  $\alpha$ -quantile of  $Y$ . We recall that for  $\alpha \in [0, 1]$  the superquantile/CVaR risk measure  $\mathcal{R}(Y) = \int_\alpha^1 G_Y(\beta) d\beta / (1 - \alpha)$ ; see Appendix. For this risk measure, it is straightforward to show that the residual risk of  $Y$  with respect to  $X$  takes the simple form

$$\mathcal{R}(Y|X) = \mu_Y + \sigma_Y \sqrt{1 - \rho^2} \frac{\varphi(\Phi^{-1}(\alpha))}{1 - \alpha},$$

where  $\varphi$  and  $\Phi$  are the probability density and cumulative distribution functions of a standard normal random variable, respectively. The value of  $c$  that attains the minimum in item (i) of Theorem 3.4 is

$-\rho\sigma_Y/\sigma_X$ . We note that for  $\rho = \pm 1$  the residual risk is reduced to its minimum possible level of  $\mu_Y$ . The other extreme is attained for  $\rho = 0$ , when  $\mathcal{R}(Y|X) = \mathcal{R}(Y)$ . In view of the previously discussed hedging investor, we note that for perfectly correlated investment possibilities, “risk” can be eliminated. The sign of the correlation coefficient is immaterial as both short and long positions are allowed. In the case of independent assets, no hedging possibility exists and the investor faces the inherent risk in  $Y$ .

We next examine the case when  $Y$  is statistically independent of  $X$  in the general case. We start with terminology.

**3.5 Definition** (representation of risk identifiers) A risk identifier  $Q^Y$  at  $Y \in \mathcal{L}^2$  for a regular measure of risk will be called representable if there exists a Borel-measurable function  $h_Y : \mathbb{R} \rightarrow \mathbb{R}$ , possibly depending on  $Y$ , such that

$$Q^Y(\omega) = h_Y(Y(\omega)) \text{ for a.e. } \omega \in \Omega.$$

For first-order and second-order superquantile/CVaR risk measures there exist representable risk identifiers for all  $Y \in \mathcal{L}^2$ ; see the Appendix.

**3.6 Proposition** Suppose that  $Z, Y \in \mathcal{L}^2$  are statistically independent. If  $Q^Y$  is a representable risk identifier at  $Y$  for a regular measure of risk, then  $Q^Y$  and  $Z$  are statistically independent.

**Proof.** Since  $Q^Y$  is a representable risk identifier, there exists a  $h_Y : \mathbb{R} \rightarrow \mathbb{R}$ , Borel-measurable, such that for almost every  $\omega \in \Omega$ ,  $h_Y(Y(\omega)) = Q^Y(\omega)$ . For Borel sets  $C, D \subset \mathbb{R}$ ,

$$\begin{aligned} \mathbb{P}\{\omega \in \Omega \mid Q^Y(\omega) \in C, Z(\omega) \in D\} &= \mathbb{P}\{\omega \in \Omega \mid h_Y(Y(\omega)) \in C, Z(\omega) \in D\} \\ &= \mathbb{P}\{\omega \in \Omega \mid Y(\omega) \in h_Y^{-1}(C), Z(\omega) \in D\} \\ &= \mathbb{P}\{\omega \in \Omega \mid Y(\omega) \in h_Y^{-1}(C)\}\mathbb{P}\{\omega \in \Omega \mid Z(\omega) \in D\} \\ &= \mathbb{P}\{\omega \in \Omega \mid Q^Y(\omega) \in C\}\mathbb{P}\{\omega \in \Omega \mid Z(\omega) \in D\}, \end{aligned}$$

where the third equality follows from the fact that  $h_Y^{-1}(C)$  is a Borel set and  $Z$  and  $Y$  are independent. Consequently,  $Q^Y$  and  $Z$  are independent.  $\square$

**3.7 Theorem** (measures of residual risk under independence) Suppose that  $Y \in \mathcal{L}^2$  and  $X \in \mathcal{L}_n^2$  are statistically independent, and  $\mathcal{R}$  is a regular measure of risk with a representable risk identifier at  $Y$  and  $Y \in \text{int}(\text{dom } \mathcal{R})$ . Then,

$$\mathcal{R}(Y|X) = \mathcal{R}(Y).$$

**Proof.** By Theorem 3.4,  $\mathcal{R}(Y|X) = \inf_{c \in \mathbb{R}^n} \varphi(c)$ , where we define  $\varphi(c) = \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle)$ . Hence, it suffices to show that  $c = 0$  is an optimal solution of this problem. The assumption that  $Y \in \text{int}(\text{dom } \mathcal{R})$  ensures that  $\partial \mathcal{R}(Y)$  is nonempty and that the subdifferential formula (see for example [23, Theorem 19])

$$\partial \varphi(c) = \left\{ E[X] - E[QX] \mid Q \in \partial \mathcal{R}(Y - \langle c, X \rangle) \right\}$$

holds. Consequently, by convexity of  $\varphi$ ,  $c = 0$  minimizes  $\varphi$  if and only if  $0 \in \partial\varphi(0)$ . Since there exists a risk identifier  $Q \in \partial\mathcal{R}(Y)$  that is independent of  $X$  by Proposition 3.6, the conclusion follows by the fact that  $E[Q] = 1$  for every  $Q \in \mathcal{Q}$  and  $E[QX] = E[Q]E[X] = E[X]$  for such an independent  $Q$ .  $\square$

### 3.3 Residual Statistics and Regression

In the same manner as a statistic  $\mathcal{S}(Y)$  furnishes optimal solutions in the trade-off formulae (6) and (7), the extended notion of *residual statistic* furnishes optimal solutions in (9):

**3.8 Definition** (residual statistic) For given  $X \in \mathcal{L}_n^2$  and a regular measure of regret  $\mathcal{V}$ , we define an associated residual statistic to be the subset of  $\mathbb{R}^{n+1}$  given by

$$\mathcal{S}^0(Y|X) := \operatorname{argmin}_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \left\{ c_0 + \langle c, E[X] \rangle + \mathcal{V}(Y - [c_0 + \langle c, X \rangle]) \right\} \text{ for } Y \in \mathcal{L}^2.$$

If in addition  $\mathcal{R}$  is a corresponding measure of risk, then an associated partial residual statistic is the subset of  $\mathbb{R}^n$  given by

$$\mathcal{S}(Y|X) := \operatorname{argmin}_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle) \right\} \text{ for } Y \in \mathcal{L}^2.$$

The motivation for the terminology “partial residual statistic” becomes apparent from the following properties.

**3.9 Theorem** (residual statistic properties) Suppose that  $X \in \mathcal{L}_n^2$  and  $\mathcal{V}$ ,  $\mathcal{R}$ ,  $\mathcal{E}$ , and  $\mathcal{D}$ , are corresponding regular measures of regret, risk, error, and deviation, respectively, with statistic  $\mathcal{S}$ . Then, the residual statistic  $\mathcal{S}^0(\cdot|X)$  and partial residual statistic risk  $\mathcal{S}(\cdot|X)$  satisfy for  $Y \in \mathcal{L}^2$ :

- (i)  $\mathcal{S}^0(Y|X)$  and  $\mathcal{S}(Y|X)$  are closed and convex, and, if  $X$  is nondegenerate, then they are also nonempty.
- (ii)  $\mathcal{S}^0(Y|X)$  and  $\mathcal{S}(Y|X)$  are compact when  $\mathcal{R}(Y|X) < \infty$  and  $X$  is nondegenerate.
- (iii) If  $c \in \mathcal{S}(Y|X)$ , then  $(c_0, c) \in \mathcal{S}^0(Y|X)$  for  $c_0 \in \mathcal{S}(Y - \langle c, X \rangle)$ , whereas if  $(c_0, c) \in \mathcal{S}^0(Y|X)$ , then  $c_0 \in \mathcal{S}(Y - \langle c, X \rangle)$  and  $c \in \mathcal{S}(Y|X)$ .
- (iv) The following alternative formulae hold:

$$\mathcal{S}^0(Y|X) = \operatorname{argmin}_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \text{ and } \mathcal{S}(Y|X) = \operatorname{argmin}_{c \in \mathbb{R}^n} \mathcal{D}(Y - \langle c, X \rangle).$$

**Proof.** For Part (i), closedness and convexity are consequences of the fact that both sets are optimal solution sets of the minimization of closed and convex functions. The nonemptiness follows from Theorem 3.4(vi). For Part (ii), suppose that the sequence  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  satisfies  $(c_0^\nu, c^\nu) \in \mathcal{S}^0(Y|X)$  for all  $\nu$ . Then, an application of Lemma 3.3, with  $Y^\nu = Y$ ,  $X^\nu = X$ ,  $b^\nu = b = \inf_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])$ , implies that  $\{(c_0^\nu, c^\nu)\}_{\nu=1}^\infty$  is bounded and  $\mathcal{S}^0(Y|X)$  is therefore compact. A nearly identical argument

leads to the compactness of  $\mathcal{S}(Y|X)$ . Part (iii) follows trivially. Part (iv) is a consequence of Theorem 3.4(i).  $\square$

Generalized linear regression constructs a model  $c_0 + \langle c, X \rangle$  of  $Y$  by solving the regression problem

$$\min_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])$$

with respect to the regression coefficients  $c_0$  and  $c$ . The choice of error measure  $\mathcal{E} = \|\cdot\|_2$  recovers the classical least-squares regression technique, but numerous other choices exist. See for example [22, 26, 21], the Appendix, and the subsequent development. It is clear from Theorem 3.9(iii) that the regression coefficients can be obtained alternatively by first computing a “slope”  $c \in \mathcal{S}(Y|X)$  and then setting the intercept  $c_0 \in \mathcal{S}(Y - \langle c, X \rangle)$ , with potential computational advantages. Moreover, Theorem 3.9 shows that points furnishing the minimum value in the definition of residual risk under regret measure  $\mathcal{V}$  coincide with the regression coefficients obtained in the regression problem using the corresponding error measure  $\mathcal{E} = \mathcal{V} - E[\cdot]$ . Further connections between residual risk and regression are highlighted in the next example.

**Example 4: Entropic risk.** In expected utility theory, the utility  $\mathcal{U}(W) = E[1 - \exp(-W)]$  of “gain”  $W$  is a well-known form, which in our setting, focusing on losses instead of gains, translates into the regret  $\mathcal{V}(Y) = E[\exp(Y) - 1]$  of “loss”  $Y = -W$ . The measure of regret  $\mathcal{V}$  is regular and generates the corresponding measure of risk  $\mathcal{R}(Y) = \log E[\exp Y]$  and measure of error  $\mathcal{E}(Y) = E[\exp(Y) - Y - 1]$  by an application of Theorem 2.2. In this case, the corresponding statistic  $\mathcal{S}$  coincides with  $\mathcal{R}$ , which implies that for  $(\bar{c}_0, \bar{c}) \in \mathcal{S}^0(Y|X)$ , we have

$$\mathcal{R}(Y|X) = \langle \bar{c}, E[X] \rangle + \mathcal{R}(Y - \langle \bar{c}, X \rangle) \text{ and } \bar{c}_0 \in \mathcal{S}(Y - \langle \bar{c}, X \rangle) = \{\mathcal{R}(Y - \langle \bar{c}, X \rangle)\}.$$

Hence,

$$\mathcal{R}(Y|X) = \bar{c}_0 + \langle \bar{c}, E[X] \rangle$$

and the residual risk of  $Y$  coincides with the value of the regression function  $\bar{c}_0 + \langle \bar{c}, \cdot \rangle$  at  $E[X]$  when that function is obtained by minimizing the corresponding error measure  $\mathcal{E}(Y) = E[\exp(Y) - Y - 1]$ .

The residual risk is directly tied to the “fit” in the regression as developed next. In least-squares regression, the *coefficient of determination* for the model  $c_0 + \langle c, \cdot \rangle$  is given by

$$R_{LS}^2(c_0, c) = 1 - \frac{E[(Y - [c_0 + \langle c, X \rangle])^2]}{E[(Y - E[Y])^2]} \quad (10)$$

and provides a means for assessing the goodness-of-fit. Although the coefficient cannot be relied on exclusively, it provides an indication of the goodness of fit that is easily extended to the context of generalized regression using the insight of risk quadrangles. From Example 1’ in [26], we know that the numerator in (10) is the mean-squared error measure applied to  $Y - [c_0 + \langle c, X \rangle]$  and the denominator is the “classical” deviation measure  $\mathcal{D}(Y) = E[(Y - E[Y])^2]$ . Moreover, the minimization of that

mean-squared error of  $Y - [c_0 + \langle c, X \rangle]$  results in the least-squares regression coefficients. According to [26], these error and deviation measures are parts of a risk quadrangle and yield the expectation as its statistic. The Appendix provides further details for the essentially equivalent case involving square-roots of the above quantities. These observations motivate the following definition of a *generalized coefficient of determination* for regression with error measure  $\mathcal{E}$  (see [21, 17] for the cases of quantile and superquantile regression).

**3.10 Definition** (generalized coefficients of determination) *For a regular measure of error and corresponding measure of deviation, the generalized coefficient of determination is given by<sup>5</sup>*

$$R^2(c_0, c) := 1 - \frac{\mathcal{E}(Y - [c_0 + \langle c, X \rangle])}{\mathcal{D}(Y)} \text{ for } c_0 \in \mathbb{R}, c \in \mathbb{R}^n,$$

and the fitted coefficient of determination is given by

$$\hat{R}^2 := 1 - \frac{\inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle])}{\mathcal{D}(Y)}. \quad (11)$$

As in the classical case, higher values of  $R^2$  are better, at least in some sense. Indeed, a regression problem aims to minimize the error of  $Y - [c_0 + \langle c, X \rangle]$  by wisely selecting the regression coefficients  $(c_0, c)$  and thereby also maximizes  $R^2$ . The error is normalized with the overall ‘‘nonconstancy’’ in  $Y$  as measured by its deviation measure to more easily allow for comparison of coefficients of determination across data sets.

**3.11 Proposition** (properties of generalized coefficients of determination) *The generalized and fitted coefficients of determination satisfy*

$$R^2(c_0, c) \leq \hat{R}^2 \leq 1 \text{ for } c_0 \in \mathbb{R} \text{ and } c \in \mathbb{R}^n; \text{ and } \hat{R}^2 \geq 0.$$

**Proof.** The upper bound follows directly from the nonnegativity of error and deviation measures. Due to the minimization in the fitted coefficient of determination,  $R^2(c_0, c) \leq \hat{R}^2$ . The lower bound is a consequence of the fact that

$$\inf_{c_0 \in \mathbb{R}, c \in \mathbb{R}^n} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \leq \inf_{c_0 \in \mathbb{R}} \mathcal{E}(Y - c_0) = \mathcal{D}(Y),$$

which completes the proof. □

The connection with residual risk is given next.

**3.12 Theorem** (residual risk in terms of coefficient of determination) *The measure of residual risk associated with regular measures of error  $\mathcal{E}$  and deviation  $\mathcal{D}$  satisfies*

$$\mathcal{R}(Y|X) = E[Y] + \mathcal{D}(Y)(1 - \hat{R}^2),$$

where  $\hat{R}^2$  is the associated fitted coefficient of determination given by (11).

---

<sup>5</sup>Here,  $\infty/\infty$  and  $0/0$  are interpreted as 1.

**Proof.** Direction application of (11) and Theorem 3.4(i) yield the conclusion.  $\square$

We recall from Theorem 2.2(i) that  $\mathcal{R}(Y) = E[Y] + \mathcal{D}(Y)$ . Theorem 3.12 shows that the residual risk is less than that quantity by an amount related to the goodness-of-fit of the regression curve obtained by minimizing the corresponding error measure.

## 4 Surrogate Estimation

As eluded to in Section 1, applications might demand an approximation of a random variable  $Y$  in terms of a better known random vector  $X$ . Restricting the attention to affine functions  $f(X) = c_0 + \langle c, X \rangle$  of  $X$ , the goal becomes how to best select  $c_0 \in \mathbb{R}$  and  $c \in \mathbb{R}^n$  such that  $c_0 + \langle c, X \rangle$  is a reasonable surrogate estimate of  $Y$ . Of course, this task is closely related to the regression problem of the previous section. Here, we focus on the ability of surrogate estimates to generate approximations of risk. In this section, we develop “best” risk-tuned surrogate estimates and show how they are intimately connected with measures of residual risk. We also discuss surrogate estimation in the context of incomplete information, often the setting of primary interest in practice.

### 4.1 Risk Tuning

Suppose that  $\mathcal{R}$  is a regular measure of risk and  $Y \in \mathcal{L}^2$  is a random variable to be approximated. Then, for a random vector  $X \in \mathcal{L}_n^2$  and  $c \in \mathbb{R}^n$ ,

$$\begin{aligned}\mathcal{R}(Y) &= \mathcal{R}\left(E[Y] + \langle c, X - E[X] \rangle + Y - E[Y] - \langle c, X - E[X] \rangle\right) \\ &\leq \lambda \mathcal{R}\left(\frac{1}{\lambda}(E[Y] + \langle c, X - E[X] \rangle)\right) + (1 - \lambda)\mathcal{R}\left(\frac{1}{1 - \lambda}(Y - E[Y] - \langle c, X - E[X] \rangle)\right),\end{aligned}$$

for all  $\lambda \in (0, 1)$  because convexity holds. Consequently, an upper bound on the one-sided difference between risk  $\mathcal{R}(Y)$  and the risk of the (scaled) surrogate estimate  $c_0 + \langle c, X \rangle$ , with  $c_0 = E[Y - \langle c, X \rangle]$ , is given by

$$\mathcal{R}(Y) - \lambda \mathcal{R}\left(\frac{1}{\lambda}(c_0 + \langle c, X \rangle)\right) \leq \langle c, E[X] \rangle + (1 - \lambda)\mathcal{R}\left(\frac{1}{1 - \lambda}(Y - \langle c, X \rangle)\right) - E[Y].$$

The upper bounding right-hand side is nonnegative because  $\mathcal{R}(Z) \geq E[Z]$  for any  $Z \in \mathcal{L}^2$  and is minimized by selecting  $c \in \mathcal{S}(Y/(1-\lambda)|X/(1-\lambda))$ . (We recall that  $\mathcal{S}(Y|X)$  is nonempty by Theorem 3.9 when  $X$  is nondegenerate.) The minimum value is the (scaled) residual risk  $(1-\lambda)\mathcal{R}(Y/(1-\lambda)|X/(1-\lambda))$  minus  $E[Y]$ . Again, in view of Theorem 3.9, such  $c$  is achieved by carrying out generalized regression, minimizing the corresponding measure of error. This insight proves the next result, which, in part, is also implicit in [22] where no connection with residual risk is revealed and positively homogeneity is assumed.

**4.1 Theorem** (surrogate estimation) *For a given  $X \in \mathcal{L}_n^2$ , suppose that  $\mathcal{R}$  is a regular measure of risk, and  $\mathcal{R}(\cdot|X)$  and  $\mathcal{S}(\cdot|X)$  are the associated measure of residual risk and partial residual statistic,*

respectively. For any  $\lambda \in (0, 1)$ , let  $Y_\lambda = Y/(1 - \lambda)$  and  $X_\lambda = X/(1 - \lambda)$ . Then, the surrogate estimate  $\bar{c}_0 + \langle \bar{c}, X \rangle$  of  $Y$  given by

$$\bar{c} \in \mathcal{S}(Y_\lambda | X_\lambda) \text{ and } \bar{c}_0 = E[Y - \langle \bar{c}, X \rangle]$$

satisfies

$$\mathcal{R}(Y) - \lambda \mathcal{R}\left(\frac{1}{\lambda}(\bar{c}_0 + \langle \bar{c}, X \rangle)\right) \leq (1 - \lambda)\mathcal{R}(Y_\lambda | X_\lambda) - E[Y]. \quad (12)$$

The surrogate estimate  $\bar{\bar{c}}_0 + \langle \bar{c}, X \rangle$ , with  $\bar{\bar{c}}_0 = (1 - \lambda)\mathcal{R}(Y_\lambda - \langle \bar{c}, X_\lambda \rangle)$ , satisfies

$$\mathcal{R}(Y) \leq \lambda \mathcal{R}\left(\frac{1}{\lambda}(\bar{\bar{c}}_0 - \langle \bar{c}, X \rangle)\right).$$

**Proof.** The first result follows by the arguments prior to the theorem. The second result is a consequence of moving the right-hand side term of (12) to the left-hand side and incorporating that term into the constant  $\bar{c}_0$ , which is permitted because  $\mathcal{R}(Y + k) = \mathcal{R}(Y) + k$  for  $Y \in \mathcal{L}^2$  and  $k \in \mathbb{R}$ .  $\square$

The positive homogeneity of  $\mathcal{R}$  allows us to simplify the above statements.

**4.2 Corollary** For a given  $X \in \mathcal{L}_n^2$ , suppose that  $\mathcal{R}$  is a positively homogeneous regular measure of risk, and  $\mathcal{R}(\cdot | X)$  and  $\mathcal{S}(\cdot | X)$  are the associated measure of residual risk and partial residual statistic, respectively. Then, the surrogate estimate  $\bar{c}_0 + \langle \bar{c}, X \rangle$  of  $Y$  given by

$$\bar{c} \in \mathcal{S}(Y | X) \text{ and } \bar{c}_0 = E[Y - \langle \bar{c}, X \rangle]$$

satisfies

$$\mathcal{R}(Y) - \mathcal{R}(\bar{c}_0 + \langle \bar{c}, X \rangle) \leq \mathcal{R}(Y | X) - E[Y].$$

The surrogate estimate  $\bar{\bar{c}}_0 + \langle \bar{c}, X \rangle$ , with  $\bar{\bar{c}}_0 = \mathcal{R}(Y - \langle \bar{c}, X \rangle)$ , satisfies

$$\mathcal{R}(Y) \leq \mathcal{R}(\bar{\bar{c}}_0 - \langle \bar{c}, X \rangle).$$

**Example 5. Risk-tuned Gaussian approximation.** Theorem 4.1 supports the construction of risk-tuned Gaussian approximations of a random variable  $Y$ , which can be achieved by considering a Gaussian random vector  $X$ . Observations of  $(Y, X)$  could be the basis for generalized regression with a measure of error corresponding to  $\mathcal{R}$ , which then would establish  $\bar{c}$  and subsequent  $\bar{\bar{c}}_0$ . Then,  $\bar{\bar{c}}_0 + \langle \bar{c}, X \rangle$  is a risk-tuned Gaussian approximation of  $Y$ . If  $\mathcal{R}$  is positively homogeneous, then  $\mathcal{R}(\bar{\bar{c}}_0 + \langle \bar{c}, X \rangle)$  is an approximate upper bound on  $\mathcal{R}(Y)$ , with the imprecision following from the passing to an empirical measure generated by the observations of  $(Y, X)$ . We next discuss such approximations in further detail.

## 4.2 Approximate Random Variables

Surrogate estimation and generalized regression are often carried out in the context of incomplete (distributional) information about the underlying random variables. A justification for utilizing *approximate* random variables is provided by the next two results. The first result establishes consistency in generalized regression and the second proves that surrogate estimates using approximate random variables remain conservative in the limit as the approximation vanishes. We refer to [30] for consistency of sample-average approximations in risk minimization problems.

**4.3 Theorem** (consistency of residual statistic and regression) Suppose that  $\mathcal{V}$  is a finite regular measure of regret and that  $Y^\nu \in \mathcal{L}^2, X^\nu = (X_1^\nu, \dots, X_n^\nu) \in \mathcal{L}_n^2, \nu = 0, 1, 2, \dots$ , satisfy

$$Y^\nu \rightarrow Y^0 \text{ and } X_i^\nu \rightarrow X_i^0 \text{ for all } i, \text{ as } \nu \rightarrow \infty.$$

If  $S^0(\cdot|X^\nu)$  are the associated residual statistics, then<sup>6</sup>

$$\limsup_{\nu \rightarrow \infty} S^0(Y^\nu|X^\nu) \subset S^0(Y^0|X^0).$$

**Proof.** Let  $c_0 \in \mathbb{R}$  and  $c \in \mathbb{R}^n$ . Since  $\mathcal{V}$  is finite, closed, and convex, it is continuous. Moreover,  $E[X^\nu] \rightarrow E[X^0]$ . For  $\nu = 0, 1, 2, \dots$ , let  $\varphi^\nu : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  be defined by

$$\varphi^\nu(c_0, c) = c_0 + \langle c, E[X^\nu] \rangle + \mathcal{V}(Y^\nu - [c_0 + \langle c, X^\nu \rangle]).$$

Then, as  $\nu \rightarrow \infty$ ,  $\varphi^\nu(c_0, c) \rightarrow \varphi^0(c_0, c)$ . Thus, the finite and convex functions  $\varphi^\nu$  converge pointwise on  $\mathbb{R}^{n+1}$  to  $\varphi^0$ , and therefore they also epiconverge to the same limit by [27, Theorem 7.17]. The conclusion is then a consequence of [27, Theorem 7.31].  $\square$

The theorem establishes that solutions of approximate generalized regression problems are indeed approximations of solutions of the actual regression problem. We observe that if  $(Y^\nu, X^\nu)$  converges in distribution to  $(Y^0, X^0)$  as well as  $E[(Y^\nu)^2] \rightarrow E[(Y^0)^2]$  and  $E[(X_i^\nu)^2] \rightarrow E[(X_i^0)^2]$  for all  $i$ , then the  $\mathcal{L}^2$ -convergence assumption of the theorem holds.

Approximations in surrogate estimation are addressed next.

**4.4 Theorem** (surrogate estimation under approximations) Suppose that  $\mathcal{R}$  is a regular measure of risk and  $\mathcal{S}(\cdot|X)$  and  $\mathcal{R}(\cdot|X)$ ,  $X \in \mathcal{L}_n^2$ , are the associated measure of residual risk and partial residual statistic. Let  $Y^\nu \in \mathcal{L}^2, X^\nu = (X_1^\nu, \dots, X_n^\nu) \in \mathcal{L}_n^2, \nu = 0, 1, 2, \dots$ , satisfy

$$Y^\nu \rightarrow Y^0 \text{ and } X_i^\nu \rightarrow X_i^0 \text{ for all } i, \text{ as } \nu \rightarrow \infty.$$

Moreover, suppose that the functional  $(Y, X) \mapsto \mathcal{R}(Y|X)$  is continuous at  $(Y^0, X^0)$ ,  $\mathcal{R}$  is continuous at 0, and  $X^0$  is nondegenerate. Then, the surrogate estimates  $\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle, \nu = 1, 2, \dots$ , of  $Y^0$  given by

$$\bar{c}^\nu \in \mathcal{S}(Y_\lambda^\nu|X_\lambda^\nu) \text{ and } \bar{c}_0^\nu = (1 - \lambda)\mathcal{R}(Y_\lambda^\nu - \langle \bar{c}^\nu, X_\lambda^\nu \rangle),$$

with  $\lambda \in (0, 1)$ ,  $Y_\lambda^\nu = Y^\nu/(1 - \lambda)$ , and  $X_\lambda^\nu = X^\nu/(1 - \lambda)$ , satisfy

$$\mathcal{R}(Y^0) \leq \liminf_{\nu \rightarrow \infty} \mu \lambda \mathcal{R}\left(\frac{1}{\mu \lambda} (\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle)\right)$$

for all  $\mu \in (0, 1)$ .

---

<sup>6</sup>Recall that for a sequence of sets  $\{A^\nu\}_{\nu=1}^\infty$ , the outer limit  $\limsup_{\nu} A^\nu$  is the collection of all points that are limits of subsequences of points selected from  $\{A^\nu\}_{\nu=1}^\infty$ .

**Proof.** Since

$$\bar{c}_0^\nu + \langle \bar{c}^\nu, X^\nu \rangle = \bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle + \langle \bar{c}^\nu, X^\nu - X^0 \rangle,$$

convexity of  $\mathcal{R}$  and Theorem 4.1, applied for every  $\nu$ , imply that

$$\mathcal{R}(Y^\nu) \leq \lambda \mathcal{R}\left(\frac{1}{\lambda}(\bar{c}_0^\nu + \langle \bar{c}^\nu, X^\nu \rangle)\right) \leq \mu \lambda \mathcal{R}\left(\frac{1}{\mu \lambda}(\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle)\right) + (1-\mu) \lambda \mathcal{R}\left(\frac{1}{(1-\mu)\lambda} \langle \bar{c}^\nu, X^\nu - X^0 \rangle\right). \quad (13)$$

Next, we establish the boundedness of  $\{\bar{c}^\nu\}_{\nu=1}^\infty$ . An application of Lemma 3.3, with  $(c_0^\nu, \bar{c}^\nu) \in \mathcal{S}^0(Y^\nu | X^\nu)$ , the associated residual statistic,  $b^\nu = \mathcal{R}(Y^\nu | X^\nu) - E[Y^\nu]$ , and  $b = \mathcal{R}(Y^0 | X^0) - E[Y^0]$  so that  $\mathcal{E}(Y^\nu - [c_0^\nu + \langle \bar{c}^\nu, X^\nu \rangle]) = \mathcal{R}(Y^\nu | X^\nu) - E[Y^\nu]$  and  $b^\nu \rightarrow b$ , implies the boundedness of  $\{(c_0^\nu, \bar{c}^\nu)\}_{\nu=1}^\infty$  and therefore also of  $\{\bar{c}^\nu\}_{\nu=1}^\infty$ . The boundedness of  $\{\bar{c}^\nu\}_{\nu=1}^\infty$  and the fact that  $X_i^\nu \rightarrow X_i^0$  for  $i = 1, \dots, n$ , result in  $\langle \bar{c}^\nu, X^\nu - X^0 \rangle \rightarrow 0$ . Since  $\mathcal{R}$  is continuous at 0, we have that  $\mathcal{R}(\langle \bar{c}^\nu, X^\nu - X^0 \rangle) \rightarrow \mathcal{R}(0) = 0$  and due to closedness,  $\liminf_\nu \mathcal{R}(Y^\nu) \geq \mathcal{R}(Y^0)$ . The conclusion therefore follows by taking limits on both sides of (13).  $\square$

Again, the positively homogeneous case results in simplified expressions.

**4.5 Corollary** *If the assumptions of Theorem 4.4 hold and the surrogate estimates  $\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle$ ,  $\nu = 1, 2, \dots$ , of  $Y^0$  are given by*

$$\bar{c}^\nu \in \mathcal{S}(Y^\nu | X^\nu) \text{ and } \bar{c}_0^\nu = \mathcal{R}(Y^\nu - \langle \bar{c}^\nu, X^\nu \rangle).$$

Then,

$$\mathcal{R}(Y^0) \leq \liminf_{\nu \rightarrow \infty} \mathcal{R}(\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle).$$

Theorem 4.4 supports surrogate estimation in the following context. Historical data, viewed as observations of an unknown random variable  $Y^0$  and a random vector  $X^0$ , can be utilized in generalized regression using an error measure corresponding to a risk measure of interest. This yields the “slope”  $\bar{c}^\nu$  and an “intercept”  $\bar{c}_0^\nu$  subsequently computed as specified in Theorem 4.4. Suppose then that the random vector  $X^0$  becomes available, for example due to additional information arriving. This is the typical case in factor models in finance where  $Y^0$  is a stock’s random return and  $X^0$  might be macroeconomic factors such as interest rates and GDP growth. Forecasts of such factors are then used for  $X^0$ . Alternatively,  $X^0$  might have been available from the beginning, which is the case when it is an input vector to a discrete-event simulation selected by the analyst. Regardless of the circumstances, the surrogate estimate  $\bar{c}_0^\nu + \langle \bar{c}^\nu, X^0 \rangle$  then provides an approximation of  $Y^0$  that is “tuned” to the risk measure of interest. If the initial data is large, then, in view of Theorem 4.4, we expect the risk of the surrogate estimate to be an approximate upper bound on the risk of  $Y^0$ .

A situation for which the mapping  $(Y, X) \mapsto \mathcal{R}(Y | X)$  is continuous, as required by Theorem 4.4, is stated next.

**4.6 Proposition** *The functional  $(Y, X) \mapsto \mathcal{R}(Y | X)$  on  $\mathcal{L}_{n+1}^2$ , given in terms of a regular measure of risk  $\mathcal{R}$ , is*

- (i) convex,
- (ii) closed at points  $(Y, X)$  where  $X$  is nondegenerate, and
- (iii) continuous if  $\mathcal{R}$  is finite.

**Proof.** Part (i) follows by a similar argument to the one leading to the convexity of  $\mathcal{R}(\cdot|X)$  for fixed  $X$ ; see Theorem 3.4. For Part (ii), we consider  $Y^\nu \rightarrow Y$ ,  $X^\nu \rightarrow X$ ,  $(c_0^\nu, c^\nu) \in \operatorname{argmin}_{c_0, c} \mathcal{E}(Y^\nu - [c_0 + \langle c, X^\nu \rangle])$ , which is nonempty due to Theorem 3.4 under the nondegenerate assumption on  $X$ , and  $\mathcal{E}(Y^\nu - [c_0^\nu + \langle c, X^\nu \rangle]) \leq b \in \mathbb{R}$  for all  $\nu$ . Hence,  $\mathcal{R}(Y^\nu|X^\nu) - E[Y^\nu] = \mathcal{E}(Y^\nu - [c_0^\nu + \langle c^\nu, X^\nu \rangle]) \leq b$  for all  $\nu$ . An application of Lemma 3.3 implies that there exist  $c_0^* \in \mathbb{R}$ ,  $c^* \in \mathbb{R}^n$ , and a subsequence  $\{(c_\nu^\nu, c^\nu)\}_{\nu \in \mathcal{N}}$ , with  $(c_0^\nu, c^\nu) \xrightarrow{\mathcal{N}} (c_0^*, c^*)$ , and  $\mathcal{E}(Y - [c_0^* + \langle c^*, X \rangle]) \leq b$ . Consequently,  $\mathcal{R}(Y|X) - E[Y] = \min_{c_0, c} \mathcal{E}(Y - [c_0 + \langle c, X \rangle]) \leq \mathcal{E}(Y - [c_0^* + \langle c^*, X \rangle]) \leq b$ , which establishes the closedness of  $\mathcal{R}(\cdot|\cdot) - E[\cdot]$  at points  $(Y, X)$  with  $X$  nondegenerate. The expectation functional is finite and continuous on  $\mathcal{L}^2$  so the closedness of  $\mathcal{R}(\cdot|\cdot)$  is also established at such points. In Part (iii) we first consider for  $c \in \mathbb{R}^n$  the functional

$$(Y, X) \mapsto \varphi_c(Y, X) := \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle),$$

which is convex and closed on  $\mathcal{L}_{n+1}^2$  by the regularity of  $\mathcal{R}$ . Since  $\mathcal{R}$  is finite,  $\varphi_c$  is also finite and therefore continuous. Thus,  $\varphi_c$  is bounded above on a neighborhood of any point in  $\mathcal{L}_{n+1}^2$ . Since  $\mathcal{R}(\cdot|\cdot) \leq \varphi_c(\cdot, \cdot)$  for all  $c \in \mathbb{R}$ ,  $\mathcal{R}(\cdot|\cdot)$  is also bounded above on a neighborhood of any point in  $\mathcal{L}_{n+1}^2$ . In view of [23, Theorem 8], the convexity and finiteness of  $\mathcal{R}(\cdot|\cdot)$  together with this boundedness property imply that  $\mathcal{R}(\cdot|\cdot)$  is continuous.  $\square$

## 5 Tracking of Conditional Values

Applications often direct the interest not only to a random variable  $Y$ , but also to random variables representing values of  $Y$  given certain realizations of a related random vector  $X$ . In particular, this is the case when the random vector  $X$  is, at least eventually, under the control of a decision maker.

We consider the situation where for  $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  and random vectors  $X \in \mathcal{L}_n^2$  and  $V \in \mathcal{L}_m^2$ , the random variable  $Y \in \mathcal{L}^2$  is of the form

$$Y = g(X, V),$$

where the equality holds almost surely<sup>7</sup>. Then, the parameterized random variables

$$Y_x = g(x, V), x \in \mathbb{R}^n,$$

represent ‘‘conditional’’ values of  $Y$ . The goal might be to *track* a specific statistic of  $Y_x$  as  $x$  varies or to select  $x \in \mathbb{R}^n$  such that  $Y_x$  is in some sense minimized or adequately low, for example as quantified

---

<sup>7</sup>Of course, conditions on  $g$  are needed to ensure that the random variable is in  $\mathcal{L}^2$ .

by the risk of  $Y_x$ . If the distribution of  $Y_x$  is unknown and costly to approximate, especially in view of the set of values of  $x$  that needs to be considered, it might be desirable to develop an approximation

$$c_0 + \langle c, x \rangle \approx \mathcal{R}(Y_x), \quad x \in \mathbb{R}^n.$$

We refer to such approximations of the risk of conditional random variables as *risk tracking*.

As indicated in Section 1, the area of statistical regression indeed examines models of conditional random variables, but typically at the level of expectations, such as in classical least-squares regression, and quantiles. We here consider more general statistics, make connections with measures of risk, and examine risk tracking. We start with tracking of statistics.

## 5.1 Statistic Tracking

We say that a regression function  $c_0 + \langle c, \cdot \rangle$ , computed by minimizing a regular measure of error, i.e.,  $(c_0, c) \in \mathcal{S}^0(Y|X)$ , *tracks the corresponding statistic* if

$$c_0 + \langle c, x \rangle \in \mathcal{S}(Y_x) \text{ for } x \in \mathbb{R}^n.$$

Of course, this is what we have learned to expect in linear least-squares regression where the measure of error is  $\mathcal{E} = \|\cdot\|_2$  and the statistic is the expectation and in this case surely a singleton. In view of the Regression Theorem in [26], this can also be counted on in situations with error measures of the “expectation type.” However, tracking might fail if the conditional statistic is not captured by the family of regression functions under consideration and even other times too as shown in [21].

The next result deals with a standard model in regression analysis, under which statistic tracking is achieved for regular error measures.

**5.1 Theorem** (statistic tracking in regression) *For given  $c_0^* \in \mathbb{R}$ ,  $c^* \in \mathbb{R}^n$ , suppose that*

$$Y(\omega) = c_0^* + \langle c^*, X(\omega) \rangle + \varepsilon(\omega) \text{ for all } \omega \in \Omega,$$

*with  $\varepsilon \in \mathcal{L}^2$  independent of  $X_i \in \mathcal{L}^2$ ,  $i = 1, \dots, n$ . Moreover, let  $\mathcal{E}$  be a regular measure of error and  $\mathcal{R}$ ,  $\mathcal{S}$ , and  $\mathcal{S}(\cdot|X)$  be the corresponding risk measure, statistic, and partial residual statistic, respectively. If  $\mathcal{R}$  has a representable risk identifier at  $\varepsilon$  and  $\varepsilon \in \text{int}(\text{dom } \mathcal{R})$ , then  $c^* \in \mathcal{S}(Y|X)$  and*

$$\hat{c}_0 + \langle c^*, x \rangle \in \mathcal{S}(Y_x) \text{ for all } x \in \mathbb{R}^n \text{ and } \hat{c}_0 \in \mathcal{S}(Y - \langle c^*, X \rangle).$$

**Proof.** Let  $\varphi : \mathbb{R}^n \rightarrow [0, \infty]$  be defined by  $\varphi(c) = \mathcal{D}(\langle c, X \rangle + \varepsilon)$  for  $c \in \mathbb{R}^n$ . In view of [23, Theorem 19] and the fact that  $\mathcal{D}(Z) = \mathcal{R}(Z) - E[Z]$ , we obtain the subdifferential formula

$$\partial\varphi(c) = \left\{ E[(Q - 1)X] \mid Q \in \partial\mathcal{R}(\langle c, X \rangle + \varepsilon) \right\}.$$

Since there exists a  $Q \in \partial\mathcal{R}(\varepsilon)$  that is independent of  $X$  by Proposition 3.6 and  $E[Q] = 1$  for every  $Q \in \mathcal{Q}$ , we have that  $0 \in \varphi(0)$  and  $c = 0$  minimizes  $\varphi$ . Moreover,  $c = c^*$  minimizes  $\mathcal{D}(\langle c^* - c, X \rangle + \varepsilon)$  and also  $\mathcal{D}(Y - \langle c, X \rangle)$ . Thus,  $c^* \in \mathcal{S}(Y|X)$  by Theorem 3.9. Finally,

$$\hat{c}_0 \in \mathcal{S}(Y - \langle c^*, X \rangle) = \mathcal{S}(\varepsilon + c_0^*) = \mathcal{S}(\varepsilon) + \{c_0^*\}.$$

Since

$$\mathcal{S}(Y_x) = \mathcal{S}(c_0^* + \langle c^*, x \rangle + \varepsilon) = \{c_0^* + \langle c^*, x \rangle\} + \mathcal{S}(\varepsilon),$$

the conclusion follows.  $\square$

**Example 6: Risk tracking of superquantile/CVaR.** Superquantile regression [21] involves minimizing the regular measure of error

$$\mathcal{E}(Y) = \frac{1}{1-\alpha} \int_0^1 \max\{0, \bar{G}_Y(\beta)\} d\beta - E[Y]$$

for  $\alpha \in [0, 1)$ , where  $\bar{G}_Y(\beta)$  is the  $\beta$ -superquantile of  $Y$ , i.e., the CVaR of  $Y$  at level  $\beta$ . The statistic corresponding to this measure of error is a superquantile/CVaR; see [25, 21] and Appendix. (We note that the risk measure corresponding to this error measure is the *second-order* superquantile risk measure, which is finite and also has a representable risk identifier; see the Appendix.) Consequently, Theorem 5.1 establishes that under the assumption about  $Y$ , there exists  $(c_0, c) \in \mathcal{S}^0(Y|X)$ , the associated residual statistic of  $\mathcal{E}$ , such that

$$c_0 + \langle c, x \rangle = \bar{G}_{Y_x}(\alpha) = \text{superquantile-risk/CVaR of } Y_x, \text{ for } x \in \mathbb{R}^n.$$

In summary, risk tracking of superquantile-risk/CVaR is achieved by carrying out superquantile regression; see [5] for an alternative approach to tracking CVaR.

## 5.2 Risk Tracking

In the previous subsection we established conditions under which generalized regression using a specific measure of error tracks the corresponding statistic. Even though one can make connections between statistics and measures of risk, as indicated in the preceding example, a direct approach to risk tracking is also beneficial. We next develop such an approach that relies on fewer assumptions about the form of  $Y$  as a function of  $X$ . The relaxed conditions require us to limit the study to conservative risk tracking.

The goal is to select  $x$  such that  $\mathcal{R}(Y_x)$  is minimized or sufficiently small for a given choice of risk measure  $\mathcal{R}$ . This is the common setting of risk-averse stochastic programming. Here, in contrast to the previous sections, there is no probability distribution associated with “ $x$ .” Still, when  $g$  is costly to evaluate, it might be desirable to develop an approximation of  $\mathcal{R}(Y_x)$ ,  $x \in \mathbb{R}^n$  through regression based on observations  $\{x^j, y^j\}_{j=1}^\nu$ , where  $x^j \in \mathbb{R}^n$  and  $y^j = g(x^j, v^j)$ , with  $v^j$  being a realization of  $V$ ,  $j = 1, \dots, \nu$ . One cannot expect that a regression function  $c_0 + \langle c, \cdot \rangle$  obtained from these observations using an error measure corresponding to a specific risk measure generally tracks  $\mathcal{R}(Y_x)$ ,  $x \in \mathbb{R}^n$ , even if sampling errors are ignored. In fact, one can only hope to track the statistic as laid out in the previous subsection. The next result, however, shows that one can achieve conservative risk tracking under general assumptions.

**5.2 Theorem** (conservative risk tracking) Suppose that  $X \in \mathcal{L}_n^2$ ,  $V \in \mathcal{L}_m^2$ , and  $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  satisfy  $g(X, V) \in \mathcal{L}^2$ ,  $g(x, V) \in \mathcal{L}^2$  for all  $x \in \mathbb{R}^n$ , and there exists an  $L : \mathbb{R}^m \rightarrow \mathbb{R}$ , with  $L(V) \in \mathcal{L}^2$ , such that

$$|g(x, v) - g(x', v)| \leq L(v)\|x - x'\| \text{ for all } x, x' \in \mathbb{R}^n \text{ and } v \in \mathbb{R}^m.$$

Let  $\mathcal{S}(\cdot|X)$  be a partial residual statistic associated with a positively homogeneous, monotonic, and regular measure of risk  $\mathcal{R}$ . If  $\bar{c} \in \mathcal{S}(g(X, V)|X)$  and  $\bar{c}_0 = \mathcal{R}(g(X, V) - \langle \bar{c}, X \rangle)$ , then for  $x \in \mathbb{R}^n$ ,

$$\mathcal{R}(g(x, V)) \leq \bar{c}_0 + \langle \bar{c}, x \rangle + \mathcal{R}(\langle \bar{c}, X - x \rangle) + \mathcal{R}(L(V)\|X - x\|) \leq \bar{c}_0 + \langle \bar{c}, x \rangle + \rho \mathcal{R}(\|X - x\|),$$

where<sup>8</sup>  $\rho = \sup L(V) + \|\bar{c}\|$ .

Moreover, the upper bound on  $\mathcal{R}(g(x, V))$  is tight in the sense that if  $\mathcal{R}$  is finite,  $\rho < \infty$ , and  $X^\nu \in \mathcal{L}_n^2$  is such that  $X^\nu \rightarrow x$ , then for  $\bar{c}^\nu \in \mathcal{S}(g(X^\nu, V)|X^\nu)$  and  $\bar{c}_0^\nu = \mathcal{R}(g(X^\nu, V) - \langle \bar{c}^\nu, X^\nu \rangle)$ ,

$$\mathcal{R}(g(x, V)) = \lim_{\nu \rightarrow \infty} \left[ \bar{c}_0^\nu + \langle \bar{c}^\nu, x \rangle + \rho \mathcal{R}(\|X^\nu - x\|) \right]$$

when  $\{\bar{c}^\nu\}_{\nu=1}^\infty$  is bounded.

**Proof.** The Lipschitz property for  $g(\cdot, v)$  implies that

$$g(x, V) \leq g(X, V) + L(V)\|X - x\| \text{ a.s.}$$

Since  $\mathcal{R}$  is monotonic as well as sublinear, we obtain that

$$\mathcal{R}(g(x, V)) \leq \mathcal{R}(g(X, V)) + \mathcal{R}(L(V)\|X - x\|). \quad (14)$$

Since

$$\bar{c}_0 + \langle \bar{c}, X \rangle = \bar{c}_0 + \langle \bar{c}, x \rangle + \langle \bar{c}, X - x \rangle,$$

sublinearity of  $\mathcal{R}$  implies that

$$\mathcal{R}(\bar{c}_0 + \langle \bar{c}, X \rangle) \leq \bar{c}_0 + \langle \bar{c}, x \rangle + \mathcal{R}(\langle \bar{c}, X - x \rangle).$$

By Corollary 4.2,

$$\mathcal{R}(g(X, V)) \leq \mathcal{R}(\bar{c}_0 + \langle \bar{c}, X \rangle).$$

Combining this result with (14) yields the first inequality of the theorem. The second inequality is reached after realizing that the monotonicity and positive homogeneity of  $\mathcal{R}$  imply that  $\mathcal{R}(\langle \bar{c}, X - x \rangle) \leq \|\bar{c}\| \mathcal{R}(\|X - x\|)$  and  $\mathcal{R}(L(V)\|X - x\|) \leq \sup L(V) \mathcal{R}(\|X - x\|)$ .

We next consider the final assertion. Since  $\mathcal{R}$  is continuous and  $\|X^\nu - x\| \rightarrow 0$ ,  $\rho \mathcal{R}(\|X^\nu - x\|) \rightarrow \rho \mathcal{R}(0) = 0$ . Moreover,

$$\bar{c}_0^\nu + \langle \bar{c}^\nu, x \rangle \leq \mathcal{R}(g(X^\nu, V)) + \mathcal{R}(\langle \bar{c}^\nu, x - X^\nu \rangle).$$

The Lipschitz property ensures that  $g(X^\nu, V) \rightarrow g(x, V)$  and the boundedness of  $\{\bar{c}^\nu\}_{\nu=1}^\infty$  results in  $\langle \bar{c}^\nu, x - X^\nu \rangle \rightarrow 0$ . In view of the continuity of  $\mathcal{R}$  the conclusion follows.  $\square$

Theorem 5.2 shows that an upper bound on the risk of a parameterized random variable can be achieved by carrying out generalized regression with respect to a *constructed* random vector  $X$ . We recall that in the setting of a parametrized random variable  $Y_x = g(x, V)$  there is no intrinsic probability distribution associated with “ $x$ .” However, an analyst can *select* a random vector  $X$ , carry

---

<sup>8</sup>Here the essential supremum is denoted by “sup.”

out generalized regression to obtain  $\bar{c}$ , and compute  $\bar{c}_0$ . The obtained model  $\bar{c}_0 + \langle \bar{c}, \cdot \rangle$  might not be conservative. However, an additional term  $\rho\mathcal{R}(\|X - \cdot\|)$  shifts the model sufficiently higher to ensure conservativeness.

The additional term  $\rho\mathcal{R}(\|X - \cdot\|)$  has an interesting form that guides the construction of  $X$ . If the focus is on  $x \in \mathbb{R}^n$  near  $\hat{x} \in \mathbb{R}^n$ , say within a “trust region” framework, then  $X$  should be nearly the constant  $X = \hat{x}$  such that  $\|X - \hat{x}\|$  is low as quantified by  $\mathcal{R}$ . We then expect a relatively low upper bound on  $\mathcal{R}(g(x, V))$  for  $x$  near  $\hat{x}$ . In fact, this situation is addressed in the last part of the theorem. However, as  $x$  moves away from  $\hat{x}$ , then the “penalty”  $\rho\mathcal{R}(\|X - x\|)$  increases.

A possible approach for minimizing  $\mathcal{R}(g(\cdot, V))$ , relying on Theorem 5.2, would be to use in generalized regression the observations  $\{x^j, y^j\}_{j=1}^\nu$ , where  $x^j \in \mathbb{R}^n$ ,  $y^j = g(x^j, v^j)$ , and realizations  $v^j$  of  $V$ ,  $j = 1, \dots, \nu$ , and a carefully selected distribution on  $\{x^j\}_{j=1}^\nu$ , centered near a current best solution  $\hat{x}$ , to construct  $\bar{c}$  and  $\bar{c}_0$  as stipulated in Theorem 5.2. The upper-bounding model  $\bar{c}_0 + \langle \bar{c}, \cdot \rangle + \rho\mathcal{R}(\|X - \cdot\|)$  could then be minimized leading to a new “best solution”  $\hat{x}$ . The process could be repeated, possibly with an updated set of observations. Within such a framework, the term  $\rho\mathcal{R}(\|X - \cdot\|)$  can be viewed as a *regularization* of the affine model obtained through regression.

The minimization of the upper-bounding model amounts to a specific risk minimization problem. In the particular case of the superquantile/CVaR risk measure at level  $\alpha \in [0, 1)$  and realizations  $\{x^j\}_{j=1}^\nu$ , with probabilities  $\{p^j\}_{j=1}^\nu$ , the minimization of that model is equivalent to the second-order cone program:

$$\begin{aligned} \min \quad & \langle \bar{c}/\rho, x \rangle + z_0 + \frac{1}{1-\alpha} \sum_{j=1}^\nu p^j z_j \\ & \|x^j - x\| - z_0 \leq z_j, \quad j = 1, \dots, \nu \\ & 0 \leq z_j, \quad j = 1, \dots, \nu \\ & x \in \mathbb{R}^n, z \in \mathbb{R}^{\nu+1}. \end{aligned}$$

We observe that the constant  $\bar{c}_0$  does not influence the optimal solutions of the upper-bounding model and is therefore left out.

## 6 Duality and Robustness

Conjugate duality theory links risk measures to risk envelopes as specified in (1). As we see in this section, parallel connections emerge for measures of residual risk that also lead to new distributionally robust optimization models.

### 6.1 Duality of Residual Risk

Dual expressions for residual risk are available from that of the underlying measure of risk.

**6.1 Theorem** (dual expression of residual risk) Suppose that  $X \in \mathcal{L}_n^2$  and  $\mathcal{R}(\cdot|X)$  is a measure of residual risk associated with a finite regular measure of risk  $\mathcal{R}$ , with conjugate  $\mathcal{R}^*$ . Then,

$$\mathcal{R}(Y|X) = \sup_{Q \in \mathcal{Q}} \left\{ E[QY] - \mathcal{R}^*(Q) \mid E[QX] = E[X] \right\} \text{ for } Y \in \mathcal{L}^2.$$

**Proof.** Let  $Y \in \mathcal{L}^2$  and  $X \in \mathcal{L}_n^2$  be fixed. We start by constructing a perturbation. Let  $\mathcal{F} : \mathbb{R}^n \times \mathcal{L}^2 \rightarrow \mathbb{R}$  be given by

$$\mathcal{F}(c, U) = \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle - U) \text{ for } c \in \mathbb{R}^n, U \in \mathcal{L}^2,$$

which is convex and also finite because  $\mathcal{R}$  is finite by assumption, and let  $U \mapsto \varphi(U) := \inf_{c \in \mathbb{R}^n} \mathcal{F}(c, U)$  be the associated optimal value function. Clearly,  $\mathcal{R}(Y|X) = \varphi(0)$  by Theorem 3.4(i). Since  $\mathcal{F}$  is finite (and also closed and convex), the functional  $U \mapsto \mathcal{F}(0, U)$  is continuous and, in particular, bounded above on a neighborhood of 0. By [23, Theorem 18] it follows that  $\varphi$  is also bounded above on a neighborhood of 0.

Next, we consider the Lagrangian  $\mathcal{K} : \mathbb{R}^n \times \mathcal{L}^2 \rightarrow [-\infty, \infty)$  given by

$$\mathcal{K}(c, Q) = \inf_{U \in \mathcal{L}^2} \left\{ \mathcal{F}(c, U) + E[QU] \right\}, \text{ for } c \in \mathbb{R}^n, Q \in \mathcal{L}^2,$$

and the perturbed dual function  $\mathcal{G} : \mathcal{L}^2 \times \mathbb{R}^n \rightarrow [-\infty, \infty)$  given by

$$\mathcal{G}(Q, v) = \inf_{c \in \mathbb{R}^n} \mathcal{K}(c, Q) - \langle c, v \rangle \text{ for } Q \in \mathcal{L}^2, v \in \mathbb{R}^n.$$

Then, the associated optimal value function of the dual problem is  $v \mapsto \gamma(v) := \sup_{Q \in \mathcal{L}^2} \mathcal{G}(Q, v)$ . By [23, Theorem 17] it follows that  $\varphi(0) = \gamma(0)$  because  $\varphi$  is bounded above on a neighborhood of 0. The conclusion then follows by writing out an expression for  $\gamma(0)$ . Specifically,

$$\begin{aligned} \mathcal{G}(Q, 0) &= \inf_{c \in \mathbb{R}^n} \mathcal{K}(c, Q) \\ &= \inf_{c \in \mathbb{R}^n} \left\{ \inf_{U \in \mathcal{L}^2} \left\{ \mathcal{F}(c, U) + E[QU] \right\} \right\} \\ &= \inf_{c \in \mathbb{R}^n} \left\{ \inf_{U \in \mathcal{L}^2} \left\{ \langle c, E[X] \rangle + \mathcal{R}(Y - \langle c, X \rangle - U) + E[QU] \right\} \right\} \\ &= \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle - \sup_{U \in \mathcal{L}^2} \left\{ E[Q(-U)] - \mathcal{R}(Y - \langle c, X \rangle - U) \right\} \right\} \\ &= \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + E[Q(Y - \langle c, X \rangle)] - \sup_{U \in \mathcal{L}^2} \left\{ E[QU] - \mathcal{R}(U) \right\} \right\} \\ &= \inf_{c \in \mathbb{R}^n} \left\{ \langle c, E[X] \rangle + E[Q(Y - \langle c, X \rangle)] - \mathcal{R}^*(Q) \right\} \\ &= \inf_{c \in \mathbb{R}^n} \left\{ E[QY] - \mathcal{R}^*(Q) + \langle c, E[X] - E[QX] \rangle \right\} \\ &= E[Q(Y)] - \mathcal{R}^*(Q) \text{ if } E[X] = E[QX], \text{ and } \mathcal{G}(Q, 0) = -\infty \text{ otherwise,} \end{aligned}$$

which results in the given formula.  $\square$

The restriction of  $\mathcal{Q}$  by the condition  $E[QX] = EQ$  is naturally interpreted as another ‘‘risk envelope.’’

**6.2 Definition** (residual risk envelope) For given  $X \in \mathcal{L}_n^2$  and risk envelope  $\mathcal{Q}$ , the associated residual risk envelope is defined as  $\mathcal{Q}(X) = \{Q \in \mathcal{Q} \mid E[QX] = E[X]\}$ .

Clearly, the subset  $\{Q \in \mathcal{Q} \mid E[QX] = E[X]\}$  of a risk envelope  $\mathcal{Q}$  is nonempty due to the fact that  $1 \in \mathcal{Q}$ ; see for example [26]. Consequently,  $\mathcal{Q}(X)$  is a nonempty convex set, which is also closed if  $\mathcal{Q}$  is closed. The discussion of this “reduced” set in the context of stochastic ambiguity is the next topic.

## 6.2 Distributionally Robust Models

We again return to the situation examined in Section 5.2 where the focus is on the parameterized random variable  $Y_x = g(x, V)$  defined in terms of a function  $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ , with  $V \in \mathcal{L}_m^2$ . We now, however, show that measures of residual risk give rise to a new class of distributionally robust optimization models capturing decisions under ambiguity.

A risk-neutral decision maker might aim to select an  $x \in \mathbb{R}^n$  such that the expected value of  $Y_x$  is minimized, possibly also considering various constraints. If risk averse, she might instead want to minimize the risk of  $Y_x$  as quantified by a regular measure of risk. It is well known that the second problem might also arise for a risk-neutral decision maker under distributional uncertainty. In fact, for every positively homogeneous, monotonic, and regular measure of risk, the dual expression  $\mathcal{R}(Y) = \sup_{Q \in \mathcal{Q}} E[QY]$  can be interpreted as computing the worst-case expectation of  $Y$  over a set of probability measures induced by  $\mathcal{Q}$ ; see for example [3, 16, 19, 2, 18] for extensive discussions of such optimization under stochastic ambiguity.

It is clear from Theorem 3.4 that the parameterized random variable  $Y_x$ , assumed to be in  $\mathcal{L}^2$  for all  $x \in \mathbb{R}^n$ , satisfies

$$E[Y_x] \leq \mathcal{R}(Y_x|V) \leq \mathcal{R}(Y_x) \text{ for every } x \in \mathbb{R}^n.$$

Here, we have shifted from  $X$  to  $V$  as the random vector that might help explain the primary random variable of interest  $Y_x$ . In this setting,  $x$  is simply a parametrization of that variable. We show next that the problem of minimizing the residual risk, i.e., solving

$$\min_{x \in \mathbb{R}^n} \mathcal{R}(Y_x; V), \tag{15}$$

leads to a position *between* the distributional certainty in the expectation-minimization model and the distributional robustness of a risk minimization model.

In view of Theorem 6.1, we see that when  $Y_x \in \mathcal{L}^2$ ,  $x \in \mathbb{R}^n$ ,  $V \in \mathcal{L}_m^2$ , and  $\mathcal{R}(\cdot; V)$  is a measure of residual risk associated with a positively homogeneous, finite, and regular measure of risk, the problem (15) is equivalent to

$$\min_{x \in \mathbb{R}^n} \left\{ \sup_{Q \in \mathcal{Q}} \{E[QY_x] \mid E[QV] = E[V]\} \right\}. \tag{16}$$

Here, the supremum is taken over a smaller set than in the case of the distributionally robust model of minimizing the risk of  $Y_x$ . In fact, the supremum is taken over the residual risk envelope  $\mathcal{Q}(V)$ . The reduction is achieved in a particular manner, which is most easily understood when the risk measure is monotonic: We recall that then  $\mathcal{R}(Y_x) = \sup_{Q \in \mathcal{Q}} E[QY_x]$  is the expected cost of  $Y_x$  for a decision maker

that only nominally believes the probability measure  $\mathbb{P}$  and considers a “worst-case” probability measure as characterized by the risk envelope  $\mathcal{Q}$ . In contrast,  $\mathcal{R}(Y_x|V) = \sup_{Q \in \mathcal{Q}} \{E[QY_x] \mid E[QV] = E[V]\}$  is the worst-case expected cost for the decision maker if she is willing to believe that the nominal probability measure  $\mathbb{P}$  at least assigns the *correct mean* to  $V$ , i.e.,  $E[V] = E_{\mathbb{P}'}[V]$ , with  $\mathbb{P}'$  being the “true” probability measure on  $\Omega$ . Of course,  $V$  can be artificially augmented to include terms like  $V_i^2$  and even random variables that do not enter  $g$  and therefore do not influence  $Y_x$  directly. Consequently,

minimizing residual risk is equivalent to minimizing a distributionally robust model under moment matching.

In view of Theorem 3.4, the solution of (15) benefits from the representation of residual risk in terms of the associated measure of regret  $\mathcal{V}$  and therefore amounts to solving

$$\min_{x \in \mathbb{R}^n, c_0 \in \mathbb{R}, c \in \mathbb{R}^m} \left\{ c_0 + \langle c, E[V] \rangle + \mathcal{V}(g(x, V) - [c_0 + \langle c, V \rangle]) \right\}, \quad (17)$$

which is convex if  $g$  is linear in its first argument or if  $g$  is convex in its first argument and  $\mathcal{V}$  is monotonic. Hence, residual risk gives rise to a tractable class of distributionally robust optimization models that captures ambiguity about the underlying probability measure.

## Appendix

Three risk quadrangles are especially relevant due to their connections with known regression techniques; see [26, 24, 25] for details:

**Example 7: Mean risk quadrangle.** For  $\lambda > 0$ , the choice  $\mathcal{R}(Y) = E[Y] + \lambda\sigma(Y)$ , where  $\sigma(Y) := \sqrt{E[(Y - E[Y])^2]}$ , is a positively homogeneous and regular measure of risk. The corresponding risk envelope  $\mathcal{Q} = \{Q = 1 + \lambda Z \mid \sqrt{E[Z^2]} \leq 1, E[Z] = 0\}$ , the regret  $\mathcal{V}(Y) = E[Y] + \lambda\sqrt{E[Y^2]}$ , the deviation  $\mathcal{D}(Y) = \lambda\sigma(Y)$ , the error  $\mathcal{E}(Y) = \lambda\sqrt{E[Y^2]}$ , and the statistic  $\mathcal{S}(Y) = \{E[Y]\}$ , which of course corresponds to least-squares regression.

**Example 8: Quantile risk quadrangle.** We recall that the  $\alpha$ -quantile,  $\alpha \in (0, 1)$ , of a random variable  $Y$  is  $G_Y(\alpha) := \min\{y \mid F_Y(y) \geq \alpha\}$ , where  $F_Y$  is the cumulative distribution function of  $Y$ . The  $\alpha$ -superquantile is  $\bar{G}_Y(\alpha) := (1/(1-\alpha)) \int_\alpha^1 G_Y(\beta) d\beta$ . The measure of risk  $\mathcal{R}(X) = \bar{G}_Y(\alpha)$  is positively homogeneous, monotonic, and regular, and gives the superquantile-risk/CVaR for  $\alpha \in (0, 1)$ . The risk envelope  $\mathcal{Q} = \{Q \in \mathcal{L}^2 \mid 0 \leq Q \leq 1/(1-\alpha), E[Q] = 1\}$ , the regret  $\mathcal{V}(Y) = E[\max\{0, Y\}]/(1-\alpha)$ , the deviation  $\mathcal{D}(Y) = \bar{G}_Y(\alpha) - E[Y]$ , the error  $\mathcal{E}(Y) = E[\max\{0, Y\}]/(1-\alpha) - E[Y]$ , and the statistic  $\mathcal{S}(Y) = [G_Y(\alpha), G_Y^+(\alpha)]$ , where  $G_Y^+(\alpha)$  is the right-continuous companion of  $G_Y(\alpha)$  defined by  $G_Y^+(\alpha) := \inf\{y \mid F_Y(y) > \alpha\}$ . Quantile regression relies on this error measure.

**Example 9: Superquantile risk quadrangle.** The second-order  $\alpha$ -superquantile  $\bar{\bar{G}}_Y(\alpha) := 1/(1-\alpha) \int_\alpha^1 \bar{G}_Y(\beta) d\beta$ , for  $\alpha \in [0, 1)$  and the choice  $\mathcal{R}(Y) = \bar{\bar{G}}_Y(\alpha)$  is a positively homogeneous, monotonic,

and regular measure of risk. The risk envelope is

$$\mathcal{Q} = \text{cl} \left\{ Q \in \mathcal{L}^2 \mid Q = \frac{1}{1-\alpha} \int_{\alpha}^1 q(\beta) d\beta, q \text{ an integrable selection from } \mathcal{Q}_{\beta}, \beta \in [\alpha, 1] \right\},$$

where  $\text{cl}$  denotes closure and  $\mathcal{Q}_{\beta}$  is the risk envelope of the quantile risk quadrangle at level  $\beta$ . The regret  $\mathcal{V}(Y) = 1/(1-\alpha) \int_0^1 \max\{0, \bar{G}_Y(\beta)\} d\beta$ , the deviation  $\mathcal{D}(Y) = 1/(1-\alpha) \int_{\alpha}^1 \bar{G}_Y(\beta) d\beta - E[Y]$ , the error  $\mathcal{E}(Y) = \frac{1}{1-\alpha} \int_0^1 \max\{0, \bar{G}_Y(\beta)\} d\beta - E[Y]$ , and the statistic  $\mathcal{S}(Y) = \{\bar{G}_Y(\alpha)\}$ . This error provides the foundation for superquantile regression [21].

The risk quadrangles of these examples, with the corresponding statistic, are summaries in Table 1; see [26] for many more examples.

	name of risk quadrangle		
functional	mean ( $\lambda > 0$ )	quantile ( $\alpha \in (0, 1)$ )	superquantile ( $\alpha \in (0, 1)$ )
statistic $\mathcal{S}$	$E[Y]$	$[G_Y(\alpha), G_Y^+(\alpha)]$	$\bar{G}_Y(\alpha)$
risk $\mathcal{R}$	$E[Y] + \lambda\sigma(Y)$	$\bar{G}_Y(\alpha)$	$\bar{\bar{G}}_Y(\alpha)$
regret $\mathcal{V}$	$E[Y] + \lambda\sqrt{E[Y^2]}$	$\frac{1}{1-\alpha}E[\max\{0, X\}]$	$\frac{1}{1-\alpha} \int_0^1 \max\{0, \bar{G}_Y(\beta)\} d\beta$
deviation $\mathcal{D}$	$\lambda\sigma(Y)$	$\bar{G}_Y(\alpha) - E[Y]$	$\bar{\bar{G}}_Y(\alpha) - E[Y]$
error $\mathcal{E}$	$\lambda\sqrt{E[Y^2]}$	$\frac{1}{1-\alpha}E[\max\{0, X\}] - E[Y]$	$\frac{1}{1-\alpha} \int_0^1 \max\{0, \bar{G}_Y(\beta)\} d\beta - E[Y]$

Table 1: Examples of risk quadrangles

We next give examples of representable risk identifiers and use the notation  $F_Y$  for the cumulative distribution function of  $Y$  and

$$F_Y^-(y) := \lim_{y' \nearrow y} F_Y(y'), \quad y \in \mathbb{R}$$

for its left-continuous ‘‘companion.’’

**Example 10: Representability of superquantile/CVaR risk identifiers.** We recall that a risk identifier  $Q^Y$  corresponding to the superquantile/CVaR risk measure  $\mathcal{R}(Y) = (1/(1-\alpha)) \int_{\alpha}^1 G_Y(\beta) d\beta$ , where  $\alpha \in (0, 1)$  and  $G_Y(\beta)$  is the  $\beta$ -quantile of  $Y$ , takes the form [25]:

$$\text{for a.e. } \omega \in \Omega, \quad Q^Y(\omega) = \begin{cases} \frac{1}{1-\alpha} & \text{if } Y(\omega) > G_Y(\alpha) \\ r_{\alpha}^Y & \text{if } Y(\omega) = G_Y(\alpha) \text{ and } F_Y(Y(\omega)) - F_Y^-(Y(\omega)) > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (18)$$

where

$$r_{\alpha}^Y := \frac{F_Y(G_Y(\alpha)) - \alpha}{(1-\alpha)(F_Y(G_Y(\alpha)) - F_Y^-(G_Y(\alpha)))}. \quad (19)$$

In this case, we set

$$h(y) = \begin{cases} \frac{1}{1-\alpha} & \text{if } y > G_Y(\alpha) \\ r_\alpha^Y & \text{if } y = G_Y(\alpha) \text{ and } F_Y(y) - F_Y^-(y) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

which is Borel-measurable. Moreover,  $h(Y(\omega)) = Q^Y(\omega)$ . Consequently, for any  $Y \in \mathcal{L}^2$ , there exists a representable risk identifier  $Q^Y$  for superquantile/CVaR risk measures.

**Example 11: Representability of second-order superquantile risk identifiers.** We find that a risk identifier  $Q^Y$  corresponding to the second-order superquantile risk measure  $\mathcal{R}(Y) = (1/(1-\alpha)) \int_\alpha^1 \bar{G}_Y(\beta) d\beta$ , where  $\alpha \in [0, 1]$  and  $\bar{G}_Y(\beta)$  is the  $\beta$ -superquantile of  $Y$ , i.e., the CVaR of  $Y$  at level  $\beta$ , takes the form [25]: for a.e.  $\omega \in \Omega$ ,

$$Q_\alpha^Y(\omega) = \begin{cases} \frac{1}{1-\alpha} \log \frac{1-\alpha}{1-F(\omega)} & \text{if } \alpha < f(\omega) = F(\omega) < 1 \\ \frac{1}{1-\alpha} \left[ \log \frac{1-\alpha}{1-f(\omega)} + 1 + \frac{1-F(\omega)}{F(\omega)-f(\omega)} \log \frac{1-F(\omega)}{1-f(\omega)} \right] & \text{if } \alpha < f(\omega) < F(\omega) \\ \frac{1}{1-\alpha} \left[ \frac{F(\omega)-\alpha}{F(\omega)-f(\omega)} + \frac{1-F(\omega)}{F(\omega)-f(\omega)} \log \frac{1-F(\omega)}{1-\alpha} \right] & \text{if } f(\omega) \leq \alpha \leq F(\omega) \text{ and } f(\omega) < F(\omega) \\ 0 & \text{otherwise,} \end{cases}$$

where  $F(\omega) = F_Y(Y(\omega))$  and  $f(\omega) = F_Y^-(Y(\omega))$ . In this case, we set

$$h(y) = \begin{cases} \frac{1}{1-\alpha} \log \frac{1-\alpha}{1-F(y)} & \text{if } \alpha < f(y) = F(y) < 1 \\ \frac{1}{1-\alpha} \left[ \log \frac{1-\alpha}{1-f(y)} + 1 + \frac{1-F(y)}{F(y)-f(y)} \log \frac{1-F(y)}{1-f(y)} \right] & \text{if } \alpha < f(y) < F(y) \\ \frac{1}{1-\alpha} \left[ \frac{F(y)-\alpha}{F(y)-f(y)} + \frac{1-F(y)}{F(y)-f(y)} \log \frac{1-F(y)}{1-\alpha} \right] & \text{if } f(y) \leq \alpha \leq F(y) \text{ and } f(y) < F(y) \\ 0 & \text{otherwise,} \end{cases}$$

where now  $F(y) = F_Y(y)$  and  $f(y) = F_Y^-(y)$ , which is Borel-measurable. Moreover,  $h(Y(\omega)) = Q^Y(\omega)$ . Consequently, for any  $Y \in \mathcal{L}^2$ , there exists a representable risk identifier  $Q^Y$  for second-order superquantile risk measures.

## References

- [1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–227, 1999.
- [2] A. Ben-Tal, D. Bertsimas, and D.B. Brown. A soft robust model for optimization under ambiguity. *Operations Research*, 58(4):1220–1234, 2010.
- [3] A. Ben-Tal and M. Teboulle. Portfolio theory for the recourse certainty equivalent maximizing investor. *Annals of Operations Research*, 31:479–500, 1991.
- [4] Z. Cai and X. Wang. Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, 147(1):120–130, 2008.

- [5] S. Y. Chun, A. Shapiro, and S. Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, 60(4):739–756, 2012.
- [6] G. Conner. The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, 15:42–46, 1995.
- [7] M.S. Eldred, L.P. Swiler, and G. Tang. Mixed aleatory-epistemic uncertainty quantification with stochastic expansions and optimization-based interval estimation. *Reliability Engineering and System Safety*, 96:1092–1113, 2011.
- [8] D. Friedman, R.M. Isaac, D. James, and S. Sunder. *Risky Curves. On the empirical failure of expected utility*. Princeton University Press, Routledge, NY, 2014.
- [9] P. Hall and H. G. Muller. Order-preserving nonparametric regression, with applications to conditional distribution and quantile function estimation. *Journal of the American Statistical Association*, 98(463):598–608, 2003.
- [10] T. Hothorn, T. Kneib, and P. Buhlmann. Conditional transformation models. *Journal of the Royal Statistical Society, Series B*, 76(1):3–27, 2014.
- [11] K. Kato. Weighted Nadaraya-Watson estimation of conditional expected shortfall. *Journal of Financial Econometrics*, 10(2):265–291, 2012.
- [12] J. Knight and S. Satchell (Eds.). *Linear Factor Models in Finance*. Butterworth-Heinemann, Oxford, UK, 2005.
- [13] P. Krokhmal, M. Zabarankin, and S. Uryasev. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, 16(2):49–66, 2011.
- [14] S. H. Lee and W. Chen. A comparative study of uncertainty propagation methods for black-box-type problems. *Structural and Multidisciplinary Optimization*, 37(3):239–253, 2009.
- [15] S. Leorato, F. Peracchi, and A. V. Tanase. Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics & Data Analysis*, 56(4):768–784, 2012.
- [16] H.-J. Luthi and J. Doege. Convex risk measures for portfolio optimization and concepts of flexibility. *Mathematical Programming B*, 104(2-3):541–559, 2005.
- [17] S. I. Miranda. *Superquantile Regression: Theory, Algorithms, and Applications*. Phd dissertation, Naval Postgraduate School, 2014.
- [18] K. Natarajan, M. Sim, and J. Uichanco. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20(4):695–731, 2010.
- [19] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Optimization*, 17(4):969–996, 2006.

- [20] F. Peracchi and A.V. Tanase. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry*, 24:471–493, 2008.
- [21] R. T. Rockafellar, J. O. Royset, and S.I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, 234(1):140–154, 2014.
- [22] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Risk tuning with generalized linear regression. *Mathematics of Operations Research*, 33(3):712–729, 2008.
- [23] R.T. Rockafellar. *Conjugate Duality and Optimization*, volume 16 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, 1974.
- [24] R.T. Rockafellar and J.O. Royset. Superquantiles and their applications to risk, random variables, and regression. *INFORMS, Tutorials in Operations Research*, pages 151–167, 2013.
- [25] R.T. Rockafellar and J.O. Royset. Superquantile/CVaR risk measures: Second-order theory. *In review*, 2014.
- [26] R.T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18:33–53, 2013.
- [27] R.T. Rockafellar and R. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft*. Springer, 3rd printing-2009 edition, 1998.
- [28] A. Ruszczynski and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.
- [29] O. Scaillet. Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal*, 72:639–660, 2005.
- [30] A. Shapiro. Consistency of sample estimates of risk averse stochastic programs. *Journal of Applied Probability*, 50:533–541, 2013.